Laws of Nature

P. Mittelstaedt P.A. Weingartner

Laws of Nature

Springer

Prof. Dr. Peter Mittelstaedt
Institut für Theoretische Physik
Universität zu Köln
Zülpicher Str. 77
50937 Köln
Germany
e-mail: mitt@thp.uni-koeln.de

Prof. Dr. Paul A. Weingartner
Institut für Philosophie
Universität Salzburg
Franziskanergasse 1
5020 Salzburg
Austria
e-mail: Paul.Weingartner@sbg.ac.at

Library of Congress Control Number: 2005927735

ISBN-10 3-540-24079-9 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-24079-2 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springeronline.com

Printed in Germany

Typesetting: by the authors and TechBooks using a Springer LaTeX macro package
Cover design: Erich Kirchner, Heidelberg

Printed on acid-free paper SPIN: 11329053 55/JVG 5 4 3 2 1 0

Preface

This book is not a textbook to become acquainted with the laws of nature. An elementary knowledge about laws of nature, in particular the laws of physics, is presupposed. The book is rather intended to provide a clarification of concepts and properties of the laws of nature.

The authors would like to emphasise that this book has been developed – created – as a real teamwork. Although the chapters (and in some cases parts of the chapters) were originally written by one of the two authors, all of them were discussed thoroughly and in detail and have been revised and complemented afterwards. Even if both authors were in agreement on most of the foundational issues discussed in the book, they did not feel it necessary to balance every viewpoint. Thus some individual and personal difference or emphasis will still be recognisable from the chapters written by the different authors. In this sense the authors feel specifically responsible for the chapters as follows: Mittelstaedt for Chaps. 4, 9.3, 10, 11.2, 12, 13 and Weingartner for Chaps. 1, 2, 3, 5, 7, 8.2, 9.2, 9.4. The remaining parts are joint sections.

Most of the chapters are formulated as questions and they begin with arguments pro and contra. Then a detailed answer is proposed which contains a systematic discussion of the question. This is the respective main part of the chapter. It sometimes begins with a survey of the problem by giving some important answers to it from history (cf. Chaps. 6 and 9). However the main part of each chapter is not historical and the authors do not identify themselves with a historical position. The main part of the chapters tries to give some systematic answer to basic questions in the light of our knowledge today. The method to begin with arguments pro and contra was chosen in order to stimulate and to draw the reader's attention also to specific problems connected with the question of the chapter. Since the problems of the pros and contras are not always central they are discussed and clarified in the commentaries (answers) to the objections at the ends of the chapters; first because these commentaries presuppose what has been said in the main part of the chapter; second they are not included in the main part of the chapter in order not to distract. It has to be emphasised however that what is expressed

in the pros and contras is not the opinion of the authors. It is sometimes the opinion of other scholars as shown by quotations. The opinion of the authors is expressed in the main part of the chapters and in the commentary to the objections.

Acknowledgements

The authors want to thank Ursula Stranzinger, Eva Stieringer and Beate Neugebauer for typing parts of the manuscript. Further thanks go to Springer, especially to Professor Beiglböck and his team, for the effective cooperation.

October 2004

Peter Mittelstaedt
Paul Weingartner

Contents

Introduction

In our ordinary experience, we observe regularities: The daily sunrise, the sequence of seasons during the year, and the regular increase and decrease of the visible size of the moon. Do these observations indicate strict laws that hold rigorously and without any exception? David Hume argued that induction is not sufficient for concluding that there are strict laws behind the observed regularities. Hence, we cannot be sure that there are laws at all and our first question reads: *Are there laws of nature?* According to Hume, Kant, and many other philosophers this question cannot be answered by induction alone. Moreover, we are also confronted with the inverse question. If there are regularities that are based on strict laws that hold necessarily, may these laws be considered as genuine laws of nature? We discuss this problem with respect to the laws of logic and with respect to some laws of mathematics. Our first, still preliminary answer is that these formal and necessary laws should not be considered as laws of nature. But then we must find an answer to the main question of these first investigations: *What is a law of nature?* We discuss this kind of problem in Part I of the present book.

Instead of giving a hasty answer to the two questions mentioned, in Part II we investigate at first properties of relations that may be considered as candidates for *"laws of nature"*. We study these problems not in the greatest possible generality but we restrict our considerations in general to physics and thus to the laws of physics. There are several reasons for this restriction. First, physics is a highly developed field of science – a mature science – which no longer consists of a large collection of isolated and merely empirically confirmed rules, but of networks of multiply connected (empirically confirmed) laws which are called *theories*. This holistic structure implies that only theories can be tested empirically and not isolated relations, which means that the laws of physics have a much higher reliability than individual law-like rules in other fields of science. Second, from a reductionistic point of view, physics may be considered as basic for science in general, since the laws of astronomy, chemistry, biology etc. can, in principle be based on the laws of physics. The reason for this important observation is not the higher accuracy and reliability of the laws of physics, but the fact that physics is concerned

with the most general structures of the empirical reality, with the most abstract laws of space, time, and matter, which are also fundamental for all other fields of science.

According to these arguments, we investigate the properties of physical laws and physical theories in more detail. In particular, we discuss the interrelations between laws and invariance principles, between laws and initial conditions, and between laws and constants of nature. Furthermore, we investigate the relations between laws of nature – in the sense of laws of physics – and causal relations. Is the principle of causality a law of nature and are laws of nature necessarily causal? And what can be said about the predictability of future events by laws of nature? Finally, we discuss the important question whether there are two kinds of physical laws, dynamical laws and statistical laws, – or whether statistical laws can always be reduced to dynamical laws. In other words, is a statistical law merely an expression of an incomplete system of dynamical laws or are there in addition also irreducible, genuine statistical laws?

The detailed knowledge of properties of laws and the answers to the various questions mentioned will help us step by step to understand the meaning of the concept "law of nature". In particular, it will become clear in what sense a law of nature refers to a structure of the real world and in what sense it is an expression of our intentions and our means of cognition. The known laws of physics contain in general objective elements referring to the external reality as well as constructive and conventional components, which are induced by subjective decisions of the scientist. Only on the basis of our knowledge of this complex structure, can we hope to successfully attack our last and most ambitious question: *Why are the laws of nature valid?*

Clearly, this question must not be understood as falling back to the metaphysics and theology of the 17th century, to the justification of laws by metaphysical principles as we find in the writings of philosophers from Descartes to Leibniz and Wolff. And we are equally not interested in the naïve recourse of this question to theology as we can find it even in the work of physicists in the 20th century. The answer we are looking for is intended to be free from metaphysical speculations and based exclusively on our detailed knowledge of the complex structure and the properties of the laws of nature.

It is obvious that we can put the question for ultimate reasons only with respect to the most general and most fundamental laws. It is meaningless to ask why Faraday's law of induction, discovered in 1831, holds. The answer is trivial today, since it follows from Maxwell's equations. Most laws of physics are imbedded in "theories" and we could ask only why these theories hold. In addition, according to some contemporary attempts the well-established theories can presumably be incorporated into a unified final theory of everything. Hence, the search for rational reasons of the laws of nature must be concentrated and restricted to the most fundamental and most abstract features of physical theories. Two examples of this kind are elaborated in more detail in Part III of the present book.

Part I

What is a Law of Nature?

1

Are there Laws of Nature at All?

This question has many facets and provides many different answers. A law of nature is a law and hence a conceptual and linguistic entity, and a law of nature refers to nature, i.e. to the real world. At first glance it is not quite clear how these two aspects fit together. Here we will briefly discuss several arguments and counterarguments which can be put forward and which might serve as a smooth introduction into the problems of the present book. All details will be discussed in the following Chaps. 1, 2 and 3.

1.1 Arguments Contra (Objections)

1.1.1 Every law of nature is a representation of some structure of nature (i.e. of the real world). However, nature, or the real world, is in permanent change. But what is an accurate representation of something in change needs to be changing too. On the other hand a law is something which does not change.

Therefore: no law is a law of nature; and consequently there are no laws of nature.

1.1.2 A law is called a law of nature in so far as it represents some structure of nature (viz. some structure of the real world). Now every law is a conceptual (or linguistic) object (entity). But conceptual (or linguistic) objects (entities) are independent of the structure of nature (structure of the real world). On the other hand no law of nature (since it represents nature) is independent of nature or the real world.

Therefore: no law is a law of nature or, what follows from that: no law of nature is a law. Thus there are no laws of nature.

1.2 Argument Pro

1.2.1 Laws of nature are usually expressed by law statements. But there are lots of law statements which have been established and extensively confirmed by the natural sciences.

Therefore: there are laws of nature expressed by law statements (of the natural sciences).

1.3 Proposed Answer

In what follows we give a preliminary answer to the first question which aims at clarifying conceptual and terminological points. Further details will become clear from further questions. The following answer will be divided into three steps: We shall start with a very wide concept of law which will be illustrated by a few examples (1.3.1). Secondly, a clarification of the concept of "law of nature" will be given with the help of several distinctions (1.3.2). Thirdly, a more detailed answer will be given to the question whether there are laws of nature at all (1.3.3).

1.3.1 Wide Concept of Law

A law is a rule, order or description by which certain things (objects) and the relations among them are arranged (ruled or ordered) or described. Concerning this one might ask three questions: (1) What kinds of things are arranged (ruled or ordered)? (2) Who has invented or discovered the rule? (3) What kind of thing is the rule (law) itself? If the law in question is a juridical law, then the things ruled by it are human actions, the inventor(s) of the rule (law) are human persons (for instance the members of a parliament) and the rule (law) is a law statement expressed in some (natural or juridical) language and announced publicly (promulgated).

If the law in question is a law of logic or mathematics (say arithmetic) then the things ruled are propositions or numbers (i.e. conceptual objects or entities), the discoverer is a logician or mathematician and the law is a law statement formulated in logical or mathematical language. Finally, if the law in question is a law of nature, then the things (with their properties and relations) described or ordered by the law are things of nature, i.e. objects of the real world (universe), the inventor of the (true) law can be the creator of the universe, the discoverer(s) of the law are human persons (scientists) and the law itself is a law statement formulated in the language of some of the natural sciences.

1.3.2 Clarification of the Concept "Law of Nature"

As to the clarification of the concept "law of nature" we shall first deal with the objects described by the law and second with different meanings (concepts) of the expression "law of nature".

1.3.2.1 As to the first it is convenient to divide things, objects or entities into three categories: natural objects, concrete human artefacts and conceptual objects.[1] Examples of natural objects are protons, planets, fields of force, lakes, plants, human persons, societies, etc.

Examples of concrete human artefacts are particular computers, houses, works of art, linguistic tokens, etc. A bird's nest or an anthill will be counted as natural objects.

Examples for conceptual objects are: concepts, propositions, sets, numbers, inferences, hypothesis, laws, theories, etc. Observe that linguistic expressions (letters, words, sentences) are usually understood not as concrete artefacts, i.e. as tokens but as conceptual objects; for instance the letter "A" is understood not as a token at a certain place but as representing a class of letters of same (similar) shape. Otherwise we couldn't say that we find the same letter in some line below, i.e. "same' means "same form'. Likewise a law statement '$dx/dt = v(x,t)$' is usually understood as representing a class of expressions (formulations) of same (similar) shape with the same meaning; i.e. it is understood as a conceptual object. Further "electron microscope" is usually understood as a conceptual object, except in the case of particular concrete electron microscope in a particular research institute (which is a concrete artefact). Observe further that all concrete artefacts are built up from (consist of) natural objects.

If we use the words "objects" and "things" we want to point out that objects are not just identical with the set of properties describing them even if they constitute often that part of the object which is known to us and which enters laws. We assume that these properties have a bearer or that the real world consists of individuals with their properties which we call "things" or "objects", though we are aware that the concept of "individual" is not an absolute one. A similar view was taken by Einstein.[2]

Concerning the question now which kinds of objects, together with their properties and relations, are described or ordered by laws of nature we can answer: natural objects and concrete artefacts.

On the other hand conceptual objects without material basis are not described or ordered by laws of nature. They are ordered and described by laws of logic and mathematics. But nothing hinders that natural objects and concrete artefacts are described by laws of nature with the help of conceptual objects.

1.3.2.2 As to the different meanings of the expression "law of nature" we notice that this expression can mean at least five different things:

L1 The "law" as it "is" in the thought of the inventor or discoverer;
L2 The "law" as it "is" in the things which are ordered or described by it;

[1] This distinction is due to Bunge (1973, MMM), p. 114.
[2] cf. Einstein (1944, BRE). See also Chap. 10.

L3 The “law” as a law statement formulated in some scientific language and belonging to some scientific theory;
L4 The “law” as an ideal true law, with respect to (w.r.t.) which laws known at present in the sense of L3 are approximations;
L5 The “law” as an ideal conceptual entity more or less independent and separated from law statements.

As to L1 we can have some partial understanding if we speak for example of (Newton’s) law of gravitation in the sense that Newton thought about it or as Newton had it in (his) mind. And historians of science are sometimes providing historical or biographical evidence for this or that version of reconstructing the thoughts and views of the genius concerning such a law. But for a discussion of “law of nature” L1 is not suitable. This can be seen as follows: (i) If Newton would not have written down his law in his language, discussion about L1 could not rest on a solid basis but would be open to speculations. (ii) Newton’s thoughts can only be known via other written linguistic expressions by him anyway. (iii) The precise formulation in his scientific writing is by all means preferable to (vague) conjectures about his thoughts; i.e. L3 is preferable by all means to L1.

Concerning the inventor or creator of a lawful universe we cannot have an adequate knowledge of the thoughts of an omniscient being on laws of nature anyway.

However, we can formulate a metaphysical principle which underlies all realistic scientific investigations: The world (universe) is ordered and structured by laws. Popper called this assumption the “law of lawfulness”. To this principle we may add a second one, connected with the first, which says that there are true laws of nature:

“To assert that there exists a true law of nature may be interpreted to mean that the world is not completely chaotic but has certain structural regularities “built in”, as it were.”[3]

“Our belief that there are true natural laws is undoubtedly based in some way or other, on observed regularities.”[4]

As Popper points out, such a belief may still be justified, even if it will be difficult to point to a particular law of physics and say: this is a true law in its present formulation and interpretation. However, what can be scientifically discussed are just these present formulations of laws in their present interpretations, that is the laws in the sense of L3.

Therefore in a discussion of “law of nature” meaning L3 is to be preferred by all means over meaning L1.

As to the second meaning of “law of nature” L2 we have to notice that a law cannot exist in the things (natural objects) as a law statement (L3) or as thoughts of a thinking person (L1). Such a view would be a too direct and naïve “picture theory” of language or of mind. We are making here only

[3] Popper (1983, RAS), p. 74.
[4] Popper (1983, RAS), p. 72. cf. Schlick (1930, FEt), p. 106.

a claim of modest realism: What corresponds to a true law of nature is a structure of things (natural objects) with their properties and relations among them. One may call such a structure a "law in the things"[5] in a derived (or metaphorical) way of speaking but one should bear in mind that there are essential differences between the meanings L2 and L3:

(i) Law statements are true or false, structures of natural (real) objects are not.
(ii) Law statements can be tested, confirmed, disconfirmed, refuted, revised, etc. Neither of those can hold for structures of real objects.
(iii) Law statements can be nearer to the truth than other law statements. Structures of real objects cannot.
(iv) Law statements can contain sentence negations and can therefore express negative facts like "there is no perpetuum mobile"; but structures of real objects cannot contain sentence negations.[6]

From (i)–(iv) it is plain that law of nature in the sense of L3 must not be confused with sense L2 and moreover that neither of the two is a mapping of the other.

Concerning meaning L4 we interpret a law statement L3 – usually – as an approximation to the true law L4 in the same intended field of application. "Usually" means here that we are never sure – even in the case of the most highly confirmed laws known – that they would not have some false consequences even if they have lots of informative and interesting true consequences; and if they have some false ones they are not completely true (having only true consequences). The true law – like the "Final Theory" – is of course not known. Nothing beyond that is claimed: For example the true law need not to have the same form or structure as the law statement which is an approximation; i.e. the true law could be non-linear whereas the approximate law statement is linear.

To say it in somewhat more general terms: Most if not all of our laws and theories will have some or other consequences which are false (already known to be false today or proved by test to be false in the future). Therefore it is important that the methods of science enable us to distinguish that law (theory) which is nearer to the truth (which has more true relevant and informative consequences and less false ones) from another law (theory) which is further away from the truth in this sense.[7]

---

[5] Recent examples are what Pagels calls the "cosmic code" (cf. Pagels (1983, CSC)) or the final symmetry structure of the universe, which some people believe to be described by "string theory" (cf. Barrow (1991, TOE), Chap. 2), or what Bohm describes as "Implicate Order" (cf. Bohm (1980, WIC)) or what Weinberg calls the "symmetry group of nature" (cf. 5.3.2(1) below).

[6] cf. Weingartner (2000, BQT), Chap. 8: Are there negative facts or properties?

[7] See Weingartner (2000, BQT)), Chap. 9: Can one theory be nearer to the truth than another?

Sense L5 of law comes close to Bolzano's and Frege's view about laws of logic and mathematics and to Popper's interpretation of them by abstracting from the knowing subject with the help of his theory of the third world.[8]

It seems to us however that sense L5 is more suitable to laws of logic and mathematics than to laws of nature. Whereas law in the sense of L5 should not be abstracted or made independent from thinking or knowing of (rational) beings in general (even if from particular ones), for laws of nature it holds in addition that they cannot be completely abstracted of or made independent from any type of real and material world (universe) of which they are descriptions.

In conclusion we want to say that the expression "law of nature" as it will be used in this book – if not otherwise indicated – is law in sense of L3; i.e. "law of nature" will be understood as a law statement formulated in the language of some scientific theory.

1.3.3 Answer to the Question: Are there Laws of Nature at All?

(1) In a preliminary sense there is a straightforward answer to this question. It is this: There are laws of nature in the sense of L3. This is evident from any textbook of physics. However this answer is preliminary at this place since important questions about the properties of laws of nature have not been discussed so far. This will be done in the subsequent chapters.

(2) There is a structure of natural objects with their properties and relations among them in the real world which is described by a true law statement or described partially by an approximate true law statement. If one wants to call this structure in a more metaphorical sense "law of nature in the real things" or "law in nature" then there are also "laws of nature" in this (metaphorical) sense.

(3) There are the thoughts of the inventor and discoverers of laws of nature. Although these thoughts (L1) are not identical with the law statements (L3) they may be expressed by linguistic signs in the form of law statements (L3). In this specific sense, namely as those thoughts of discoverers which are expressed by a law statement (L3) laws of nature also exist in the sense of L1.

(4) Laws of nature in the sense of L4 do not yet exist in the thoughts of discoverers (even if they can exist in the thoughts of an omniscient being) but they can "exist" in the sense of L5 if they are not understood as completely independent from the real world (universe). Since the usual realist understanding is that it is a necessary condition that laws of nature are dependent on nature (on the real world) in so far that they are

[8] cf. Bolzano (1929, WSL)), Sects. 20–25; Frege (1964, BLA)), Introduction; Popper (1969, EKS).

descriptions of the real world and can be discovered and invented from investigating the real world.[9]

1.4 Answer (Commentary) to the Objections

1.4.1 (to 1.1.1) "Structure of the real world" can be interpreted in a twofold way: (1) First as contingent structure. To this structure belong all initial conditions, singularities, random conditions (except constants of nature) and possible microstates or branchings which make up the same macro-state. Of this contingent structure it is correct to say that it changes permanently. (2) Second as necessary and invariant structure. To this structure belong properties and relations of things (objects) of the real world (universe) which are conserved and invariant relative to a set of changes of certain magnitudes. This can be illustrated by the following examples: Consider first the simple spherical pendulum: although the pendulum is in permanent change there are invariant structures like the period of oscillation (neglecting damping) or the relation between length and period of oscillation. A further simple example is this: $p \cdot V$ (pressure times volume) is invariant with respect to the changes of the magnitudes (quantities) p and V (where T is kept constant). When T is incorporated, $p \cdot V = R \cdot T$ is a better approximation. Of this structure, in the second sense, it is not correct to say that it changes. Now laws of nature describe the invariant, necessary and conserved structure of nature (of the real world).[10] They abstract from hic et nunc (from here and now) as already Thomas Aquinas pointed out very clearly;[11] i.e. they do not tell us singularities, particular initial states, particular random or microstates. Since laws of nature describe the invariant and conserved structure of the world they do not change even if they are accurate descriptions.

Thus the answer to the objection 1.1.1 is this: The second premise of the argument uses the contingent structure of the world, whereas the fourth premise uses the conserved (invariant) structure. Since this is a fallacy of equivocation the conclusion of the argument is not proved.

1.4.2 (to 1.1.2) Concerning the independence of conceptual or linguistic objects from the structure of the real world we have to notice that two different meanings of independence have to be distinguished here: The first is concerned with the different ontological status (1) and the second is concerned with the deviation of the law statement from the correct (true) description of the world (2).

(1) As to the first we see that the means by which we describe the world do usually not have the same ontological status as the world. Only in

[9] See Sects. 2.1.1–2.1.3 and 2.4.1–2.4.3 for the view of Kant.
[10] For further details see Chaps. 5 and 6 on invariance.
[11] Thomas Aquinas (STh)I, 46,2.

exceptional cases we use things to signify things as we use a particular flag to signify a particular nation or a portrait to refer to a particular person, or we use linguistic entities to describe other linguistic entities. But in the normal case and especially in the sciences we use conceptual and linguistic entities to signify and describe things of this world.

Now this difference in ontological status does certainly imply a certain kind of independence which is quite different for different natural languages: Wittgenstein's idea of a picture theory of language in his *Tractatus* – only *one* mapping structure can be the correct picture of the real world – was given up by himself in his *Philosophical Investigations*, where he says of this view: "Ein *Bild* hielt uns gefangen."[12] The diversity of natural languages with their different structures – compare Indo-European languages with Arabic languages or with Chinese or Japanese languages – refute every too simple minded picture theory. Moreover no scientific languages, as for example the languages of modern physics or chemistry have a simple picture structure.

(2) Concerning the second meaning of independence we understand that a law statement should be *dependent* (i.e. *not* independent) on the real, or corresponding to the real world in the sense that every new knowledge (experiment) about the real world may lead to a revision and correction (even perhaps refutation) of it. If however the law statement deviates from the true law (or even from a better approximation to it) it has a certain degree of independence w.r.t. certain series of tests. Thus for example Galileo's law for throwing bodies and the parabola as the trajectory cannot be corrected (or refuted) by different arrangements of tests with stones from (even high) towers. Or the interpretation of mass as velocity-independent (expressed in the Euler formulation of Newton's second law) could not be tested with different tests using the then available velocities. This kind of independence however is only accidental in the sense that it is connected with the tests available of that time such that new and further developed tests will make the respective law statements dependent of the real world again.

Thus the answer to the objection 1.1.2 is this: The first kind of independence (difference of ontological status) does not imply the second one (independence from the real world). On the contrary: The first kind of independence is perfectly compatible with dependence in the second sense. In fact premise 3 of the argument uses "independence" in the first sense, whereas premise 4 uses "independence" in the second sense. Therefore the conclusion of 1.1.2 is not proved by this argument.

[12] Wittgenstein (1960, PhI)), §115.

2

Can the Laws of Nature be Genuine Laws?

After we will give some arguments pro and contra we shall propose an answer as follows: First we shall give eight necessary conditions for a law being a genuine law. Then we shall discuss these conditions in some detail and with examples. Finally we shall show that what is usually and often vaguely understood by *law of nature* fits very well these conditions for genuine laws.

2.1 Arguments Contra

2.1.1 As Kant says, every genuine law is universally valid and necessary. “Nun sind wir gleichwohl wirklich im Besitze einer reinen Naturwissenschaft, die a priori und mit aller derjenigen Notwendigkeit, welche zu apodiktischen Sätzen erforderlich ist, Gesetze vorträgt, unter denen die Natur steht”, “Es sind daher objektive Gültigkeit und notwendige Allgemeingültigkeit (vor jedermann) Wechselbegriffe”,[1] “Es finden sich aber unter den Grundsätzen jener allgemeinen Physik etliche, die wirklich die Allgemeinheit haben, die wir verlangen... Diese sind wirklich allgemeine Naturgesetze, die völlig a priori bestehen.”[2]

If a law is universally valid it is obeyed by every kind of entity (object); thus the laws of logic and mathematics – like for example $x = x$ – are obeyed by every entity (object), i.e. by conceptual objects, natural objects and concrete artefacts. But as it was said in 1.3.21 laws of nature are not obeyed by all kinds of objects but only by natural objects and concrete artefacts; thus they are not universally valid.

Therefore: laws of nature are not genuine laws.

2.1.2 According to Kant it holds that if a law is a synthetic a priori truth then it cannot be falsified or revised. Moreover according to Kant, every genuine law is a synthetic a priori truth. But laws of nature can be falsified or revised. This

[1] Kant (1783, PzM), Sects. 15 and 19.
[2] Kant (1783, PzM), Sect. 15.

can be demonstrated by the fact that many laws of nature have been either corrected and revised or falsified. To give only two examples: Galileo's law of falling bodies (presupposing constant acceleration) was revised by Newton because the gravitational force is dependent on height. The universal validity of the law of conservation of parity was refuted by Lee and Yang.

Therefore: laws of nature are not genuine laws.

2.1.3 What is necessary holds in every possible world. Now as Kant says every genuine law is necessary: "Da das Wort Natur schon den Begriff von Gesetzen bei sich führt, dieser aber den Begriff der Notwendigkeit"[3] But laws of nature do not hold in every possible world. Since there might be consistent (possible) worlds with different laws of nature.

Therefore: laws of nature are not genuine laws.

2.2 Arguments Pro

All laws of nature are either very general and severally confirmed dynamical laws or very general and severely confirmed statistical laws. But all dynamical or statistical laws, which are very general and severely confirmed are genuine laws. Therefore: All laws of nature are genuine laws.

2.3 Proposed Answer

Laws of nature are genuine laws. One way to show this is with the help of general conditions for genuine laws. Further support for this answer will become evident from subsequent chapters. Here we characterise genuine laws by several conditions, which should be fulfilled. The list (2.3.1) of eight conditions is not considered complete. It gives a framework that is filled by the investigations of the following chapters.

2.3.1 Genuine Law

A law, understood as a law statement in the sense of L3 (i.e. formulated in the language of some scientific theory), is a genuine law if it satisfies conditions G1–G8:

G1 Genuine laws describe or order a large domain of objects (natural objects, concrete artefacts or conceptual objects) together with their properties and relations among them. This domain belongs to some science.

G2 Genuine laws describe the conserved and invariant properties of the objects in that domain but they abstract from the objects as individuals.

[3] Kant (1786, MAN), Vorrede, p. VI.

G3 Genuine laws are invariant w.r.t. a certain set of changes of their parameters.
G4 Genuine laws are spacetime invariant.
G5 Genuine laws hold either in all cases of application (i.e. without exception) or in most cases.
G6 Genuine laws are a good approximation to the true law, i.e. they have a high degree of confirmation and informative content.
G7 Genuine laws belong to a system of laws, which makes up the central part of a theory.
G8 Genuine laws refer (ultimately) to objective reality.

2.3.2 Commentary to the Conditions G1–G8

G1: Genuine laws describe or order a large domain of objects (natural objects, concrete artefacts or conceptual objects) together with their properties and relations among them. This domain belongs to some science.

Laws usually apply to objects of a certain domain. And for genuine laws the domain has to be quite large; thus Newton's laws apply to all bodies which are moved by forces as he tells us in the Preface to his Principia:

> "I wish we could derive the rest of the phenomena of Nature by the same kind of reasoning from mechanical principles, for I am induced by many reasons to suspect that they may all depend upon certain forces by which the particles of bodies, by some causes hitherto unknown, are either mutually impelled towards one another, and cohere in regular figures, or are repelled and recede from one another."[4]

On the other hand the domain of the laws of Boyle–Mariotte and Gay-Lussac or that of Snell's refraction law is rather restricted from the very beginning. Therefore these "laws" will not be called genuine laws.

However this does not mean that they are not strictly valid in a narrower domain. Such a restriction of the domain is usually achieved by certain conditions like $T =$ constant or $V =$ constant in the case of Boyle–Mariotte and Gay-Lussac respectively; or like the laws of a pendulum or oscillator when restricted to small amplitudes, or Hooke's law for small deformations.

G1 may also be expressed by saying that a genuine law has a high degree of universality. Though it is important to observe here that the universality must not be artificial, virtual or reductionist; in any case a (several) universal quantifier(s) is not sufficient to guarantee universality: thus in $\forall x[x = c \rightarrow Px]$ the variable x is reduced to a constant c in the antecedent such that this sentence expresses a pseudo-universality.[5]

[4] Newtond (Princ), Preface, p. XVIII.

[5] This point has been emphasised by Popper already in his (Popper (1959, LSD) p. 68) by the example: For all x if x is identical with Napoleon then x is born in Corsica. It holds also for such pseudo-examples as: for all x if x is an apple in this basket then x is red. They are also ruled out as laws by condition G5.

Since laws of nature are usually understood as describing natural objects (cf. 1.3.1) or objects of nature together with their properties and relations (including concrete artefacts) they must have a very wide domain of application. Thus they satisfy G1 for genuine laws.

We may also say that the appropriate domain of laws of nature is the domain of the natural sciences, although this does not mean that every particular law would encompass this whole domain. Thus laws of nature coincide with those genuine laws of which their appropriate domain is that of the natural sciences or a subdomain of it. One may speak of genuine laws also outside this domain if all the eight necessary conditions G1–G8 are satisfied. However, laws of logic and mathematics can count as genuine laws only if condition G6 and G8 are interpreted in a very wide sense: i.e. if it is not required of the confirming and refuting instances that they represent contingent facts with concrete spacetime parameters and if "objective reality" in G8 may be interpreted also as conceptual reality in a similar sense as Popper's "world 3" (cf. Chaps. 3 and 4).

Observe that a single law of nature usually does not describe the whole domain of objects belonging to the natural sciences. This could only be possible if this law would contain the "Final Theory" or a "Theory of Everything" in the sense of a "World Formula". The whole domain of the natural sciences (natural objects and concrete artefacts) is not even completely covered by all the laws of nature we know today, though in principle they refer to that global domain. Even the most general laws we have, pick out some subdomain by special antecedence conditions which are usually not explicitly mentioned. Thus particle mechanics picks out rigid bodies measured by rigid measurement rods (which are assumed to be arbitrarily movable in space). Special relativity picks out mechanical and electromagnetic objects under Lorentz transformation without taking into account *gravitation.*

G2: Genuine laws describe the conserved and invariant properties of the objects in that domain but they abstract from the objects as individuals.

No law describes individuals as individuals, though it describes individual objects inclusively, since it describes the properties of objects of a certain *kind*; i.e. the properties of protons, neutrons, electrons, magnetic fields, electric fields, gases, etc. But to exchange two individuals (say two electrons or two protons) does not affect any law.[6] Therefore laws of nature, by which we understand very general laws, do not describe the individual as individual (i.e. as different from another individual) but are invariant with respect to an exchange of individuals. And so they satisfy condition G2 for genuine laws.

G3: Genuine laws are invariant w.r.t. a certain set of changes of their parameters.

[6] This property is called permutation invariance; for details see Chap. 5.3.2(1).

Already Aristotle observed that (his) substances are invariant w.r.t. several kinds of changes. In this sense the following passage comes close to formulating a law of substance conservation:

> "But clearly matter also is substance; for in all the opposite changes that occur there is something which underlies the changes, e.g. in respect of place that which is now here and again elsewhere, and in respect of increase that which is now of one size and again less or greater; and in respect of alteration that which is now healthy and again diseased; and similarly in respect of substance there is something that is now being generated and again being destroyed, and now underlies the process as a 'this' and again underlies it as the privation of positive character. In this last change the others are involved. But in either one or two of the others this is not involved; for it is not necessary if a thing has matter for change of place that it should also have matter for generation and destruction."[7]

Every law is an invariance condition w.r.t. certain sets of changes.[8] And for genuine laws these changes have a rather wide range. Now for genuine laws the set of changes that do not change the law is usually twofold: there is the set of changes which concerns the parameters explicitly belonging to the law itself. For example the parameters mass and distance in Newton's law of gravitation: $F = G \cdot m_1 \cdot m_2/r^2$. In addition there is the set of changes belonging to certain background conditions or presuppositions. For example the position or constant velocity of the whole system, which do not influence the relative positions and velocities of the masses inside the system. Both sets of changes have restrictions and it would not make sense to require invariance w.r.t. "all" changes. However the set of changes concerning the parameters explicitly occurring in the law must not break down for extensions (provided if they are not too extreme) otherwise it cannot be a genuine law: this is the case with the "laws" of Boyle–Mariotte, Gay-Lussac, Rayleigh–Jeans and Wien which hold only for a relatively small domain of changes. Since laws of nature are understood as general principles (of nature) which express a symmetry or invariance, it is clear that they must satisfy G3 (cf. 5.3.2 and 5.3.3). The search for the symmetry group of nature (i.e. the set of all changes which do not change laws of nature) is, according to Weinberg, the deepest thing that we understand about nature: "It is increasingly clear that the symmetry group of nature is the deepest thing that we understand about nature today... Specifying the symmetry group of nature may be all we need to say about the physical world beyond the principles of Quantum Mechanics."[9]

---

7 Aristotle (Met, 1042a 33-b6).

8 For a detailed exposition of this important property of laws see Chaps. 5 and 6

9 Weinberg (1987, TFL), p. 73.

G4: Genuine laws are spacetime invariant.

Spacetime invariance could have been included under G3. This has not been done for the following reason: The invariance with respect to displacement of time and place is the oldest and perhaps most important invariance property of physical laws and of laws of nature in general. These two invariances are of such fundamental importance that it is justified to say that the concept of law (and of genuine law and law of nature) could not be understood if spacetime invariance would not be satisfied:

> "This principle can be formulated, in the language of initial conditions as the statement that the absolute position and the absolute time are never essential initial conditions. The statement that absolute time and position are never essential initial conditions is the first and perhaps the most important theorem of invariance in physics. If it were not for it, it might have been impossible for us to discover laws of nature."[10]

G5: Genuine laws hold either in all cases of application (i.e. without exception) or in most cases.

Already Aristotle had such a principle for the sciences:

> "But that there is no science of the accidental is obvious; for all science is either of that which is always or of that which is for the most part."[11]
> "Since, among things which are, some are always in the same state and are of necessity (not necessity in the sense of compulsion but that which means the impossibility of being otherwise), and some are not of necessity nor always, but for the most part, this is the principle and this the cause of the existence of the accidental; for that which is neither always nor for the most part, we call accidental."[12]

If we translate it into modern terms it says: Science is either of that which can be described by strict or dynamical laws or of that which can be described by statistical laws. If we do not have laws of either kind we cannot do science.[13]

It should be noted that accepting G5 for genuine laws and also for laws of nature has severe consequences: i.e. to take statistical laws serious and to interpret them realistically. That means to accept branching or degrees of freedom as objectively given in nature and not just interpreted epistemically as degrees of our ignorance. Though we do not, of course, deny that independently of that, degrees of ignorance are everywhere (due to men's and scientist's imperfection).

---

[10] Wigner (1967, SRf), p. 4. For details cf. Chaps. 6 and 8

[11] Aristotle (Met), 1027a 21.

[12] Aristotle (Met), 1026b27f.

[13] This principle of Aristotle throws out already a lot of the artificial examples discussed concerning laws by philosophers like questions whether "all apples in this basket are red" is a law. cf. G1.

Therefore it is also justified to investigate our degrees of ignorance with a subjective interpretation of probability. But what we want to stress here is that there are, in our view reasons in nature (reality) for an objective interpretation which treats the degrees of freedom and branching as *real*. However this difficult problem cannot be discussed in detail here.[14]

In contradistinction to dynamical laws, there is an important exception for statistical laws from the requirement that a genuine law of nature holds in all cases of application. A statistical law does not provide predictions about properties of a single object but only about probabilities and thus about relative frequencies of the appearance of an observable property. However, a predicted relative frequency will not appear in all cases. There are – very few – sequences of events, the "non-random sequences", that do not show the predicted relative frequency behaviour, even if the sequences are sufficiently long. A violation of the relative frequency prediction is not strictly impossible but only "almost impossible". Hence a statistical law holds in "almost all" or "most" cases (cf. Chap. 7).

G6: Genuine laws are a good approximation to the true laws: i.e. they have a high degree of confirmation and informative content.

This condition G6 includes three parts. First that laws can be understood as approximations to the true law and consequently that the concept of approximation to truth can be made precise. Secondly and thirdly that necessary conditions for a good approximation to the true law are a high degree of confirmation and a high degree of informative content.

(1) As to the first we want to point out that we accept the general idea of Karl Popper's theory of "verisimilitude"[15] according to which theory (law) A is closer to the truth than theory (law) B iff A has more true and less false consequences than B.

This (or more accurately the respective precise definition) has been correctly criticised by Tichy and Miller.[16] It can be shown however that Popper's original idea can be rehabilitated if one restricts the consequence class of classical logic to non-redundant and most informative consequence elements.[17] Moreover it can be shown that this "reduced" consequence class is not weaker in logical content, i.e. is logically equivalent (by classical logic) to the full consequence class.[18]

In the revised sense of Popper's original idea we can say for example that Newton's theory of motion is nearer to the truth (his laws are nearer to the true law) than Galileo's theory. Newton's laws are a better

[14] For more on this difficult problem see Chap. 13.

[15] Popper (1963, CaR), Appendix 3 and Popper (1972, Okn), p. 330.

[16] cf. Tichy (1974, PDV), Miller (1974, CFT).

[17] This has been shown in Schurz, Weingartner (1987, VDR). cf. also Weingartner (2000, BQT), Chap. 9.

[18] Therefore one does not lose any logical strength by this reduction (in contradistinction to an erroneous remark of Niiniluoto (1998, VTP)).

approximation to the true law than Kepler's three laws. In both eases Newton's laws have more true and less false consequences than Galileo's or Kepler's: Thus Galileo's incorrect parabola is replaced by Newton's correct ellipse and it is shown (by Newton) that a trace of a projectile becomes a parabola only under the incorrect assumption that the radius of the earth is infinite such that the distance of the flight becomes negligible compared to the size of the surface on which it lands. Kepler's third law $a^3/T^2 = \text{const}$, Newton replaced by $a^3/T^2 = \text{const}\ (m_0 + m_1)$ – more accurately by $a^3/T^2 = \gamma/4\pi^2 \cdot (m_0 + m_1)$ – which is a theoretically more detailed formulation and for the general case a better approximation.[19] It should however be mentioned that $a^3/T^2 = \text{const}$ gives remarkably good results even with today's exact measurements of T and r (for the first 4 planets deviation occurs only at the fifth decimal place). In this case Newton's improvement, which is theoretically more correct for the general case, does not give better results if only two masses m_0 and m_1 are taken into account. An analogous comparison could be done also with several other important consequences of the above three theories (three sets of laws).[20] Now every law of nature satisfies G6 in this respect: only if a law can be called a good approximation to the true law it will be called a law of nature. Thus it is justified to call Newton's laws of motion or his law of gravitation a law of nature, whereas this would not be justified for Galileo's law of falling bodies.

A further point to observe is that some hypotheses, since they have been refuted soon or immediately, never were considered as approximations; whereas others, although it was discovered that they have some false consequences, are still considered as laws which are good approximations to the true law. An example for the first case is Bohr's atom model and one for the second case are Newton's laws of motion, which have been superseded by Einstein's.

A further example for progress towards a better approximation to the true law is the development from both the Rayleigh–Jeans law of radiation (false for short waves) and Wien's law of radiation (false for long waves) to Planck's law of radiation.

(2) It is a necessary condition for a genuine law that it has a high degree of confirmation, which means that it has withstood a great number of important and severe tests. Observe that neither quantity nor probability are sufficient to make a test severe: Galileo's law of falling bodies could be confirmed with an unlimited number of tests with stones thrown from high towers. The consequence is not that the quantity of tests is unimportant

[19] Here m_0, m_1 are the masses of the sun and the planet, a is the distance and T is the time of revolution, γ is Newton's gravitational constant.

[20] Such a comparison concerning approximation to the true law has been done in a precise way with the development of the gas law in Schurz, Weingartner (1987, VDR).

but that it may be empty if the test is done in such a restricted area within which the law (hypothesis) is almost trivially satisfied. According to a frequency interpretation of probability, a hypothesis which passes only every second of a series of tests will get probability $1/2$ though it will be thrown away immediately; moreover a hypothesis which passes 90 of the tests and fails 10 serious ones will also be thrown away immediately, i.e. will be considered as unconfirmed, though having a high probability w.r.t. this interpretation. This shows that a high degree of confirmation of a hypothesis (law) cannot be defined in a simple and direct way w.r.t. the quantity of positive tests. Though the frequency interpretation of probability is the best objective way to use probabilities in natural science, it is concerned with single events and does not give directly a probability measure of statements (hypothesis or laws).[21]

The above mentioned laws of Wien and Raleigh–Jeans are also examples of "laws" with a low degree of confirmation: they were never well confirmed before Planck replaced them by the much better approximation of his law of black body radiation, which has been very well confirmed then.

Also with respect to point (2) we can say that only a law with a high degree of confirmation can be called a law of nature such that also laws of nature satisfy this condition for genuine laws.

(3) A further necessary condition for a genuine law is a high degree of informative content. What is informative content is dependent on the area of research and has to be relativised respectively. Thus with respect to logical consequence informative content goes together with logical strength or logical content.

Logical content can be defined in two different ways. Inside the field of logic it is concerned with the logical strength of axioms of a system of logic (say for propositional or predicate logic): it can be defined as the class of consequences (in this case of logically true sentences) and it depends on the used derivation rules. But this is an area of proof theory and does not concern us here. Outside the field of logic, i.e. when logical inference (logical consequence) is applied to different sciences, then the logical content of a sentence A is defined as the class of all logical consequences of A (i.e. those derivable by logical deduction) which are not logically true. Thus $A \wedge B$ (A and B) has more logical content than $A \vee B$ (A or B) since from the former both A and B can be derived whereas neither of them can be derived from the latter. In this sense every genuine law has a huge logical content.

Another type of informative content which is important for all empirical sciences is the so-called *empirical content—f* "empirical content"; it

[21] This point has been stressed by Popper already in his German version (1934) of (1959, LSD), p. 201ff and 215ff and in several of his later writings. cf. also Bunge (1967, SRII), p. 323.

can be defined as the class of all possible and consistent empirical counterexamples of a hypothesis or a law. More accurately: the empirical content of a hypothesis H (or law in the sense of L3) is the class of all possible (consistent) empirical test statements which criticise (refute, falsify) H.[22]

A test statement is a consistent statement with concrete spacetime parameters which describes the result of an observation or experiment. Observe that empirical content of H could not be defined by the class of all empirical test statements which are derivable from H; because from a hypothesis or law in the sense of L3 no such statements are derivable. In order to derive an empirical test statement from a law statement we would have to add to the law statement other empirical statements with concrete parameters like initial conditions, random conditions, constants etc. The above condition that genuine laws have to have empirical content can also be expressed by saying that genuine laws are not trivial, because they forbid or they rule out a lot of possible empirical events or situations concerning natural objects. A law of logic, say $p \Rightarrow p$ or $p \wedge (p \rightarrow q) \Rightarrow q$ allows every empirical event (except logically inconsistent or impossible ones). Therefore it does not say anything specific w.r.t. empirical phenomena and consequently does not have empirical content. A genuine law however says something specific about empirical or factual events, i.e. selects certain events as factual and therefore has to forbid a lot of others which are logically possible but not factual.[23] Again another way of saying that genuine laws have empirical content is to say that they are synthetic in Kant's sense.

Finally concerning laws of nature it must be said that only a law with empirical content can be called a law of nature. This is so because a law without empirical content does not say anything specific about natural objects and their properties since it would allow any kind of object and any kind of event if they are just logically consistent.

G7: Genuine laws belong to a system of laws which makes up the central part of a theory.

This condition requires *systemicity* or belonging to some scientific system (theory) whether already developed to a full-fledged theory or yet on a lower level of developing. There are several subconditions for systemicity:

[22] The definition of "empirical content" is due to Popper (1959, LSD); included already in his Logik der Forschung of 1934) VI, Sect. 35 "I define the of a statement p as the class of its potential falsifiers". Popper uses "basic statement" instead of "test statement". (1959, LSD), Sect. 28: "Basic statements have the form of singular existential statements". Example: There is a so and so in the (spacetime) region k. For other types of informative content like mathematical content, value content and normative content cf. Weingartner (1978, WThI), p. 45f.

[23] For more on the distinction between laws of logic and mathematics on the one hand and laws of nature on the other see Chaps. 3 and 4.

(1) First that there must not be an inconsistency among the laws which belong to the same theory; i.e. inside one and the same theory the laws have to be mutually compatible. This requirement of mutual compatibility is in fact much more general and holds (at least as a normative principle though not always as a factual one) also among scientific disciplines: thus mathematical laws have to be compatible with laws of logic, physical laws with laws of mathematics, chemical and biological laws with those of physics, anthropological and sociological ones with those of biology. In this respect we accept also the requirement that philosophical theories concerning nature (natural objects) have to be compatible with well confirmed scientific theories.

(2) Secondly, since the laws make up the gist (kernel) of a theory, a change of one of the laws will effect the theory such that one would speak of a change of the theory as well. On the other hand the periphery of a theory consists of (less confirmed) hypothesis. A change in such a hypothesis will not change the theory.

(3) Thirdly a law, even if it is well confirmed but isolated, will not be called genuine law. For example take Archimedes' principle for the flotation of bodies which is one of the oldest examples of what was called a physical law. This principle was well confirmed for solid/fluid pairs. But only in the 19th century it could be incorporated into mechanics where flotation could be understood as a particular case of balancing of forces. By so belonging to a scientific theory, this principle got indirect confirmation support via the confirmation of the principles of dynamics.[24] Thus two important features can be grasped from this example: A principle or a hypothesis which is incorporated into a theory receives first a new and generalised interpretation of its basic concepts (in this case: flotation is interpreted as a particular case of balancing of forces) and secondly it receives support from the confirmation of other laws of the theory and from the confirmation of the whole theory. Further examples are the radiation laws of Rayleigh–Jeans (wrong for short waves) and of Wien (wrong for long waves) which have never been incorporated into a theory before they were superseded by Planck's radiation law which then became the start and a basic building block of quantum theory.

The following two points (4) and (5) are optional conditions, not necessary conditions:

(4) Fourthly, there may be a subordination of laws inside a theory. For instance if one law is discovered to be a less general case of a still more general law: The vector principle of the parallelogram is derivable from classical particle mechanics; but also the much more general law of the conservation of the linear impetus is a consequence from the laws of

[24] cf. the detailed and interesting discussion of the principle of Archimedes in Bunge (1967, SRI), p. 329ff.

classical particle mechanics. In the latter case we have a subordination of laws.

(5) Fifthly, there may be a special interrelation between a law and the theory in which it is incorporated: interpretability. A law L is (weakly) interpretable[25] in a theory T if there is a consistent extension T^* of T such that for every non-logical basic concept C of L there are definitions in T^* which define C with the help of basic concepts in T. Examples: Peano's number theory is interpretable in the axiomatic theories (systems) of set theory. The so-called equivalence principle – which says that it is always possible to arrange the transformations of the spatial coordinates in such a way that the static homogeneous field of gravitation will disappear – is interpretable in the laws of general relativity.

Finally it is easily seen that these conditions concerning systemicity which are characteristics of genuine laws are also characteristics of laws of nature: they have to be compatible with one another, they belong to the kernel of a theory, they are interpreted within the theory and they receive confirmation support from it; further they *might* – these are not necessary conditions – stand in a relation of deduction (derivation) to more general laws and they might be interpretable in the theory to which they belong.

G8: Genuine laws refer ultimately to objective reality.

This condition says that genuine laws formulated as law statements in the sense of L3 have as their referents patterns of objective reality. This will become more clear by contrast w.r.t. three cases in which one could not speak of genuine law:

(1) Phenomenological hypotheses. A phenomenological[26] hypothesis describes the input and output, but does not describe or explain the deeper events in-between. For example the chemical description of the photosynthesis by CO_2 + H_2O + chlorophyll + light (energy) from the sun $\Rightarrow$ glucose + oxygen is a phenomenological hypothesis.

    Only later it became known that and how the plant disintegrates water such that the hypothesis described the deeper level and became representational. A phenomenological hypothesis could never directly develop into a genuine law though a representational one may so develop.

(2) Experiment and model referent hypothesis. Hypothesis may describe series of experiments or models as approximate pictures of real systems. In both cases such hypotheses could not develop into genuine laws which have to refer to the reality "behind" such pictures or models or experimental situations.

(3) Meta-hypotheses or meta-laws. Meta-hypotheses and meta-laws refer to hypotheses and to laws. Thus "all sociological hypotheses are statistical"

[25] The concept of interpretability is due to Tarski, cf. Tarski, Mostowski, Robinson, (1968, UDT), Chap. I.

[26] cf. the description in Bunge (1967, SRI), p. 248.

and "all physical laws are (or: have to be) spacetime invariant" refer to hypotheses and laws and at least not directly to objective reality. They describe properties of laws or hypothesis or they prescribe (in the sense of norms) properties of laws or hypotheses but not properties of natural objects like physical objects or societies. Therefore – in this respect – they cannot count as genuine laws.

There is however the possibility that some laws can be formulated both in the object language about physical objects and in the meta-language about laws (as meta-laws). Take for example the principle of special relativity. Expressed in the object language about physical systems it says: All inertial systems are equivalent. Expressed in the meta-language about laws it says: The physical laws are invariant w.r.t. a transformation from one inertial system to another one. Concerning such meta-laws we may say then that not their versions about laws, but their versions about physical reality (or objective reality in general) can be called genuine laws (when satisfying the other conditions). Since laws of nature are understood as saying something about nature (i.e. natural objects and their properties) but not as saying something about laws, they behave like genuine laws: meta-laws can count as laws of nature in a derived or indirect sense.[27]

It may be worth mentioning here that there are also philosophical or metaphysical meta-laws which have some relation to laws of nature. We shall give two examples: Kant's so-called "Copernican revolution" is expressed by saying that the laws of nature are to be found in our intellect and that our intellect cannot grasp them from nature but prescribes them to nature or puts them on nature.[28] Thomas Aquinas said that the first basic laws underlying ethics are (have to be) invariant w.r.t. a transformation from one value scale (system of ethical values) to another one. Such a law is the principle: the good should be done and the bad should be avoided. This principle is invariant w.r.t. different value scales because it is formal in the sense that "good" and "bad" is open concerning its content. It is then a separate task of a special system of ethics to provide a definition or interpretation of these "variables" ("good" and "bad") and to give a reasonable justification for it.

2.4 Answer to the Arguments

2.4.1 (to 2.1.1) Genuine laws have to be universally valid and necessary in their appropriate domain, which is restricted in Kant's philosophy to objects of experience. Thus genuine laws in the area (domain) of the natural sciences need not to be obeyed by all kinds of objects generally but only by all kinds of objects of their appropriate domain. For example if the appropriate domain of physical laws are matter-dependent entities, then fictional entities of works

[27] cf. Sect. 5.3.1(3) for further details.

[28] We do however not adhere to this more idealistic view on laws of nature.

of art need not obey these laws; at least not completely, even if it is assumed that they obey such laws partially in order to be consistent with a selected part of human experience. Since the appropriate domain of laws of nature is the domain of the natural sciences they can be very well genuine laws.

2.4.2 (to 2.1.2) As it is clear from 1.3.2.2, we have to distinguish the law in the sense of L3 (i.e. law statement formulated in some scientific language and belonging to some scientific theory) from that in the sense of L4 (the ideal true law), where laws in the sense of L3 are approximations to the true law in the sense of L4. Now Kant has in mind the law in the sense of the ideal true law (L4) when he speaks of the synthetic a priori truth. Genuine laws however are usually understood as laws in the sense of L3 and because of satisfying the necessary conditions G6 and G7 they are understood as good approximations to the true law. Nothing hinders therefore that genuine laws understood in this way can be corrected and revised and sometimes even be falsified. The same is true for laws of nature. Therefore laws of nature can be very well genuine laws. And they are in fact genuine laws since they satisfy all the necessary (together sufficient) conditions for genuine laws as has been substantiated above.

2.4.3 (to 2.1.3) The word “necessary” has different meanings. The two different meanings which have to be distinguished here are “logically necessary” and “empirically necessary” or “naturally necessary”. The propositions which hold in every possible world are called logically necessary. These are the theorems of fist order predicate calculus with identity which can be substantiated by suitable axiom systems of modal logics and their semantics. Intuitively we would count to this class also the theorems of the arithmetic of natural numbers and of some other basic domains of mathematics (though these are not included in the above axiom systems and their semantics). Genuine laws, if they satisfy conditions G1–G8, are not understood as logically necessary, but as empirically or naturally necessary.[29] And this means they have empirical content and exclude a lot of possible worlds. Similarly laws of nature do of course not hold in all possible worlds, otherwise they would not exclude any possible world and thus would not say anything specific about a (or a class of) particular world(s). Therefore laws of nature can very well be genuine laws.

2.4.4 (to 2.2) The second premise in the argument has to be completed: It is not sufficient for genuine laws that they are very general dynamical or statistical laws which are severely confirmed. Also the other necessary conditions G2–G8 have to be added. The same holds for laws of nature. Then the conclusion that all laws of nature are genuine laws is satisfied.

[29] cf. Chap. 9.

3

Are the Laws of Logic Laws of Nature?

3.1 Arguments Pro

3.1.1 If the laws of logic are applicable to all objects whatsoever they are also applicable to the objects dealt with by the natural sciences. Now a law is certainly applicable to some domain of objects if all the objects of this domain satisfy the law. Thus if the laws of logic are satisfied by all objects whatsoever they are also satisfied by the objects dealt with by the natural sciences. But those laws which are satisfied by the objects dealt with by the natural sciences are called laws of nature. Consequently: if the laws of logic are applicable to all objects whatsoever then the laws of logic are also laws of nature. Now, according to Kant the laws of logic are applicable to all objects whatsoever: "Die Logik ist ... eine Wissenschaft a priori von den notwendigen Gesetzen des Denkens, aber nicht in Ansehung besonderer Gegenstände, sondern aller Gegenstände überhaupt."[1]

Therefore: the laws of logic are *also* laws of nature.

3.1.2 If logic is the most general of all sciences then its universe of discourse (UL) includes the universes of discourse of all other sciences and so also that of the natural sciences (UN), i.e. UN $\subset$ UL. It follows from this that the laws of logic which rule the elements of UL rule also the elements of UN. But the laws which rule the elements of UN are called laws of nature. Consequently if logic is the most general of all sciences then the laws of logic are laws of nature, since they rule the elements of UN. Now as Leibniz and Gödel say logic is the most general of all sciences: "Logica est scientia generalis"[2] – "it is a science prior to all others, which contains the ideas and principles underlying all sciences."[3]

Therefore: The laws of logic are *also* laws of nature.

[1] Kant (1800, Lg), A4.
[2] Leibniz (1903, OFI), p. 557.
[3] Gödel (1944, RML), p. 125.

3.1.3 If the laws of logic can be either justified by sense perception or falsified and revised by contingent facts then they are rather laws of thought or laws of thought processes and not – as Frege said – laws of truth.[4] And if they are laws of thought or thought processes they are laws of nature. Now as Gödel says of Russell the laws of logic can be justified by sense perception:

> "He [Russell] compares the axioms of logic and mathematics with the laws of nature and logical evidence with sense perception, so that the axioms need not necessarily be evident in themselves, but rather their justification lies (exactly as in physics) in the fact that they make it possible for these 'sense perceptions' to be deduced."[5]

Moreover as Quine says the laws of logic can be revised or superseded like one physical theory supersedes another:

> "Conversely, by the same token, no statement is immune to revision. Revision even of the logical law of the excluded middle has been proposed as a means of simplifying quantum mechanics; and what difference is there in principle between such a shift and the shift whereby Kepler superseded Ptolemy, or Einstein Newton, or Darwin Aristotle?"[6]

Therefore: The laws of logic are laws of nature.

3.1.4 Any formal language of science contains the most general laws of logic. Truth can be introduced w.r.t. these general laws of logic by describing their invariance against changing elementary (atomic) sentences (with the predicates contained in them) with the help of substitution: "A logical truth, then, is definable as a sentence from which we get only truths when we substitute sentences for its simple sentences."[7] Now since the laws of logic (or the logical truths) are general they can be applied to any area of science outside logic; i.e. the substitution instances can be taken from mathematics, physics, biology, etc. But such substitution instances put constraints on the laws of logic according to limits of mathematical reasoning or limits of nature. That this is so can be seen from the following two examples:

(i) As Intuitionism points out, in the case of application to an infinite domain, $p \vee \neg p$ fails because neither $\exists x Fx$ holds in its intuitionistic interpretation (there is no construction of a natural number k with a proof of Fk) nor $\neg\exists x Fx$ holds (there is no uniform proof for $\neg Fn$ for each n).
(ii) Since the principle of classical physics: *Any two properties (quantities) out of all observables can be observed (measured) simultaneously* does not hold in quantum physics, the combination or fusion of arbitrary propositions

[4] Frege (1969, NGS), p. 139.
[5] Gödel (1944, RML), p. 127.
[6] Quine (1951, LPV), p. 43.
[7] Quine (1970, PLg), p. 50.

(and properties) by any connective – though generally allowed by the laws of logic (logical truths in Quine's sense) – is limited.

But since such limitations are ruled by laws of nature, it seems that the laws of logic are limited by or reduced to laws of nature.

Therefore the laws of logic seem to be laws of nature.

3.1.5 All laws of (human) thought are laws of nature; since all humans belong to nature. Now as Boole says, the laws of logic are derived from laws of thought: "Chapter III: Derivation of the Laws of the symbols of Logic from the laws of the operations of the human mind."[8]

Therefore it seems that the laws of logic are derived from the laws of nature.

3.2 Arguments Contra

3.2.1 As Ockham and Leibniz point out the laws of logic are an instrument for all other sciences in that they are principles of demonstration and proof:

"Logica enim est omnium artium optissimum instrumentum, sine qua nulla scientia perfecte scire potest."[9]

"Scientiam generalem intelligo quae modum docet omnes alias scientias ex datis sufficientibus inveniendi et demonstrandi."[10]

But the laws of nature are not principles of demonstration and proof and consequently are not an instrument of all the sciences.

Therefore: the laws of logic are not laws of nature.

3.2.2 According to Wittgenstein the laws of logic are analytic: "The propositions of logic are tautologies. Therefore the propositions of logic say nothing. (They are analytic propositions.)"[11] But the laws of nature describe reality and are therefore not analytic but synthetic.

Therefore: the laws of logic cannot be laws of nature.

3.3 Proposed Answer

In every scientific discipline we can distinguish three domains: the domain of problems, the domain of application and its proper domain. This holds also for logic, for mathematics and for the natural sciences. But with respect to a comparison of all three domains it can be shown that the laws of logic are not laws of nature. This can be seen as follows:

[8] Boole (1854, LoT), p. 39.
[9] Ockham (1957, SLg), Pars Prima, p. 7.
[10] Leibniz (1965, GPh), VII, p. 60.
[11] Wittgenstein (1960, TLP), 6.1, 6.11. cf. 3.4.6 below.

3.3.1 The Domain of Problems

If we speak of the domain of logic we may have in mind that logic deals with a certain group of questions or problems. Although the selection of these problems is not arbitrary or just a matter of convenience it nevertheless does not have a sharp border w.r.t. other disciplines, especially to mathematics. This does not mean however that the border is not sharp everywhere, since there are also clear cases. Thus it means that there are problems which belong unambiguously either to logic or to mathematics, but that there are others where the decision is not so easy.

As to the unambiguous cases the problem of a general criterion for distinguishing valid from invalid arguments, or in other words the problem of the concept of logical consequence, will certainly belong to the domain of problems of logic. Whereas the problem of the solution of certain diophantine equations, like Fermat's problem, or the more general question of Hilbert's 10th problem, whether every diophantine equation has also a solution in integers[12] is unambiguously a problem of mathematics.

Concerning the cases where the demarcation is not sharp, Skolem points to the relation of basic concepts of logic and of arithmetic:

> "I am taking the liberty here of a remark about the relation between logical and arithmetical primitives (basic concepts). Independently whether one introduces the concept of propositional function in the first or second way one comes across the idea of the integer. Also with an axiomatic introduction of that concept (for example in investigations of consistency) one is forced to consider what can be deduced with the help of finitely many applications of the axioms. On the other hand it is not possible to logically characterise the series of numbers without using the concept of propositional function... It seems to me therefore to be ill conceived both to base the logical concepts on the arithmetical ones and vice versa. Both have to be established simultaneously and in mutual connection."[13]

Skolem's argument shows that neither all important concepts of mathematics are included in or reducible to logic – a part of the doctrine of Logicism – nor all important concepts of logic are included in or reducible to concepts of mathematics. Nevertheless there is a certain hierarchy which will be shown in the section about the proper domain of logic.

[12] The problem was solved in 1970 by Jury Matijasevic who showed that Hilbert's 10th problem is equivalent to Turing's problem which depends on the undecidability of the "halting problem". The result is roughly as follows: To a given computer programme one can construct a diophantine equation which has a solution in integers iff the computer programme comes to a stop. And conversely to a given diophantine equation one can construct a computer programme which will stop iff the given diophantine equation has a solution in integers.

[13] Skolem (1970, MLg), p. 196 (our translation).

Independently of systematic distinctions the question what belongs to the domain of problems (of a discipline) is to a great extend a question of historical development. Thus the first problem concerning the validity of arguments is as old as logic itself, i.e. goes back to Aristotle. And the question how to solve diophantine equations goes back to Diophantus, a Greek mathematician.

Concerning the demarcation of problems of logic from those of natural science it is clear that the problem to define the validity of inferences or the consequence class of a set of premises is one of logic and not one of natural science. Consequently from the point of view of the demarcation of problems laws of logic in the sense of laws (rules) of derivation or inference cannot be laws of nature. On the other hand the prize question of King Oscar II of 1885 is clearly a problem of physics (at least since mass points are not logical entities):

> "For an arbitrary system of mass points which attract each other according to Newton's laws, assuming that no two points ever collide, give the coordinates of the individual points for all time as the sum of a uniformly convergent series whose terms are made up of known functions."

The prize was given to Poincaré, who gave reasons that such series diverge rather than converge although he did not solve the problem. It was solved partially by Kolmogorov and finally by his pupil Arnold in 1963.[14]

Consequently Newton's laws and laws concerning the divergence of adjacent points cannot be laws of logic.

3.3.2 The Domain of Application

With respect to the domain of application we consider three points: (1) That the domain of application of logic with its laws is rather unlimited whereas the domain of application of natural science and its laws is narrower. (2) That the laws of logic of the full Classical Logic have different properties than the laws of nature and (3) that the laws of logic of a restricted or reduced classical calculus or those of a weaker (than classical) logic can nevertheless not be reduced to laws of nature.

(1) Unlimited domain of application
    The domain of application of logic is not only very wide, it seems rather unlimited.[15] Thus a logical inference – say the dictum de omni (inference from all cases to one particular) – can be applied to sentences of logic, of mathematics, physics, chemistry, history, juridical norms, etc. where the variables (occurring in the inference) are interpreted respectively as

[14] For more on that see Sect. 9.4.3.2(g)

[15] In a similar way is the domain of application of mathematics rather unlimited (cf. Chap. 4).

signs for numbers, elementary particles, chemical elements, negotiations for peace, human actions, mental processes.

Or the principle that if two properties (concepts) are compatible (have a common extension) then also their respective higher order properties. The analogon in set theory is: if two sets have a common member then also their supersets. Like the above inference these principles about properties and sets are also applicable to any consistent properties or sets.

But the domain of application of laws of nature is clearly narrower since we cannot say that they are applicable to logic (for example to the relations of pure concepts) to mathematics (for example to number theoretic problems) or (under the assumption of not being an extreme reductionist) to mental processes. Since no law of logic has such a restriction of application, no law of logic is a law of nature.

(2) Full classical logic (PL1)

What is usually applied both in the area of mathematics and in the sciences is much more than the above examples of very general logical rules and similar ones: It is the theorems of first order predicate logic with identity (PL1, which is also called classical logic). This system of Logic is also the underlying logic of the two systems of set theory widely used today: Zermelo–Fraenkel set theory and Neumann–Bernays–Gödel set theory. PL1 is a first order theory in the sense that there is only quantification over individual variables but not over predicates. Quantification over entities of higher type (like predicates) or over sets is done in higher order logic or in set theory. PL1 can be justified semantically or model theoretically on one hand or proof theoretically on the other. In both ways one can show that PL1 is a complete theory, but not decidable. However PL1 restricted to one-place predicates (and syllogistics which is included in it) are decidable theories. But the laws of nature (at least those which are known today) are not complete; i.e. it is not the case that all true statements about things of nature or – when restricted to physics – objects of physics or physical systems are derivable from the laws plus initial conditions.[16]

Therefore the laws of logic (as theorems of PL1) are not laws of nature (and not laws of physics).

(3) Weaker logics and restricted classical logic

(a) Weaker logics

There can be good reasons for being more modest when choosing the underlying logic for mathematics or for the application in the empirical sciences. As to mathematics Intuitionism restricts the valid laws of PL1 to those of intuitionistic logic, because it requires a special kind of definite limitation of the concepts in the mathematical propositions which are applied to the infinite domain or in other words

[16] For a detailed justification of the incompleteness of physical laws see Weingartner (1997, CLN) and Chap. 11.

of the quantified statements, where the quantification runs over an infinite domain. Observe for example that the use of the principle of excluded middle (bivalance) has never been questioned for finite domains. But the intuitionistic interpretation of the quantified form $\exists x A(x) \vee \neg\exists x A(x)$ of this principle reveals the main limitations: In order to establish the validity of $\exists x A(x)$ one must provide a construction of a natural number k and a proof of $A(k)$. To show that $\neg\exists x A(x)$ one has to give a proof which shows the falsity of $A(n)$ for all n. ($\neg\exists x A(x) \Leftrightarrow \forall x \neg A(x)$ holds intuitionistically). Thus what is required is this: Either there is a construction of a natural number k and a proof of $A(k)$ or there is a uniform proof that shows the falsity of $A(n)$ for each n. It is easily understandable that neither may be the case, i.e. the principle of excluded middle is not satisfied then.[17] As a consequence of that also the underlying Propositional Logic is affected such that bivalence and other principles like double negation, $(p \to q) \to (\neg p \vee q)$, $(\neg q \to \neg p) \to (p \to q)$, $(p \to q) \vee (q \to p)$, $\neg\neg p \to p$ are not generally valid. Moreover all the four connectives ($\wedge$, $\vee$, $\to$, $\neg$) are independent (not interdefinable).

Similar weakenings of PL1 for the applications both in mathematics and in empirical sciences are dialogical logic and constructivist logic which both have simple proof theoretical justifications.[18]

A further group of weaker logics are relevant logics and paraconsistent logics. The first group, initiated by Anderson and Belnap,[19] is based on the idea of strengthening implication which traditionally goes back even to Stoic logic and to Medieval logic but began in the 20th century with Lewis' system of "Strict Implication" (1918, SSL), Parry's "Analytische Implikation" (1933, AAI) and Ackermann's "Strenge Implikation" (1956, BSI). This approach has been further developed by Meyer, Routley, Dunn and others. Another approach rather independent of that concentrates also on implication and tries to develop a general theory of conditionals; Chellas (1975, BCL), v. Benthem (1984, FCL) and others. Paraconsistent logics concentrate on avoiding the principle $(p \wedge \neg p) \vdash q$ (ex falso quodlibet) which allows to derive any arbitrary formula from a contradiction. It was called "exploding principle" (by G. Priest) and its avoidance is a common feature of paraconsistent logics, relevant logics and restricted classical logics (CL).[20] In paraconsistent quantum logics also $A \to (B \to A)$ (which is used for the

[17] cf. Fraenkel, Bar Hillel, Levy (1973, FST), Chap. IV: Intuitionistic Conception of Mathematics (by D. van Dalen).

[18] Such proposals have been made in detail by Paul Lorenzen. cf. his (1955, EOL), (1980, MeM).

[19] cf. Anderson, Belnap (1975, Ent). For further development see Dunn (1986, RLE).

[20] See the citation of Tarski below (3.4.3). For paraconsistent logics cf. DaCosta (1974, TIF) and Priest (1987, ICd).

definition of commensurability) is invalid. This law (also called Frege's law) is also thrown out in restricted classical logic (cf. (b) P1 example 3 below).

All these logics try to solve some of the problems coming up if logic is applied to empirical sciences and they are successful to solve some of these problems, especially those connected with so-called paradoxes of implication (of which the ex falso quodlibet is one, and that truth is derivable from any premise is another). And to solve such problems was (at least originally) the motive for deviating from CL. But although the motivation came from the application to empirical sciences including natural sciences it does not follow from that that the laws of these weaker logics are laws of nature.

Another type of weaker logics are quantum logics which focus on problems arising when CL (mainly the propositional part of PL1) is applied to modern physics, especially quantum physics. CL makes two assumptions which are violated in quantum mechanics (QM):

(i) Every proposition (having a certain truth value) can be conjoined (by some connective $\wedge, \vee, \rightarrow, \leftrightarrow$) with any other proposition to result in a new proposition. Consequently also the predicates contained in the propositions can be arbitrarily combined. A consequence from this is: if proposition A represents (describes) a (measurable) physical state (or state of affairs) and proposition B does also, then the conjunction (and the disjunction, implication and equivalence) represent (describe) also a (measurable) physical state (or state of affairs). It is well known that this (kind of arbitrary combination or commensurability) is not the case in some situations (measuring position and momentum) in QM (cf. Sect. 13.2).

(ii) The laws of distributivity hold unrestrictedly as an equivalence. Already Birkhoff and v. Neumann showed in their paper (1936, LQM) that the distribution laws of CL do not hold in one direction (of the two implications) w.r.t. propositions describing experiments in QM.

Thus every quantum logic has to avoid these classical assumptions. It is known that this can be achieved with an orthomodular lattice calculus as has been shown for example by Mittelstaedt.[21] Now quantum logic is an example where facts of nature (experimental facts of QM) are in conflict with special assumptions of CL ((i), (ii) above). But observe that assumption (i) is not really a law or theorem of CL (PL1) but concerns the very liberal formation rules for building up the calculus (of PL1), although being the base for the rule of adjunction: $p,\ q \mid\!- \ p \wedge q$ (used in systems of natural deduction). And the

[21] cf. Mittelstaedt (1978, QLg). For a detailed exposition see Chap. 13 on Quantum Logic.

assumption (ii) claims an equivalence of which one part of the implication (the one which leads from disjunctive parts to conjunctive parts) still holds unrestrictedly also when applied to phenomena and experiments in QM.

What this shows is that only a part of the laws of PL1 (CL) can be used when applied to certain fields of empirical sciences, for example QM. And the other part not used here can still be used in the application to formal sciences (for example to classical mathematics and set theory). However it does not show that some of the laws of logic of PL1 (CL) are falsifyable by empirical facts as laws of nature are and therefore could be called also laws of nature.

(b) Restricted classical logic (RCL)
In contradistinction to calculi described under (a) (weaker logics) RCL is not a new logic and is in its general form in fact independent of the underlying logic or its concept of a valid inference, although this presentation uses CL as the underlying logic. RCL functions like a filter put on CL, it concentrates on restricting the classical consequence class; because it has been shown in a number of articles that redundant parts which are permitted in the consequence class of CL are the culprits of most of the difficulties and paradoxes discussed for more than 50 years in areas like confirmation, explanation, law statements, disposition predicates, epistemic and deontic logic, verisimilitude and quantum logic. Especially there are two very general properties P1 and P2 of CL (PL1) which make trouble when CL is applied outside logic and mathematics. As will be seen these two properties of CL include also the assumptions (i) and (ii) above. The properties P1 and P2 are concerned with the consequence (conclusion) α of a (classically) valid inference $A \mid\!- \alpha$.

P1 Parts of α are replaceable (in some inferences) by arbitrary parts of the same category salva validitate of the inference.

P2 Parts of α are reducible (in some inferences) to simpler (and usually shorter) parts consisting of conjuncts salva validitate of the inference and by preserving the logical content of the conclusion.

Simple examples for P1 are: $A \mid\!- (B \rightarrow A)$, $\neg A \mid\!- (A \rightarrow B)$, $A \mid\!- A \vee B$, $A \wedge \neg A \mid\!- B$, $A \mid\!- ((A \wedge B) \vee (A \wedge \neg B))$, $A \wedge B \mid\!- ((A \wedge C) \vee (B \wedge \neg C))$. In the first four valid inferences the variable "B" can be replaced by any arbitrary propositional variable salva validitate of the inference. In the fifth "B" and in the sixth "C" can be replaced on both occurrences by any arbitrary propositional variable salva validitate of the inference.[22]

Simple examples for P2 are: $C \wedge C$, $C \vee C$, $\forall x(A \wedge B)$, $A \wedge B$, $A \vee (B \wedge C)$, $(A \wedge B) \vee (A \wedge C)$ can be reduced respectively to C; $\forall xA$, $\forall xB$; A, B; $(A \vee B) \wedge (A \vee C)$; $A \wedge (B \vee C)$.

[22] We use *salva validitate* w.r.t. inferences and *salva veritate* w.r.t. statements. "$\mid\!-$" stands for derivability or valid implication.

> That P1 (replacable parts) causes difficulties[23] can be seen from the fact that replacing "B" by any other variable above means that A can be combined (by connectives) with any arbitrary proposition. Or if we think of the predicates in the propositions it means, that any predicate can be combined with any other. Physically this means that commensurability is always granted. Observe in this connection that the last two of the above examples for P1 are the propositional analoga to the definiens of commensurability and to one form of Bell's inequalities.[24]

It will be easily understood that parts (in the consequence class) which are replaceable by arbitrary parts (salva validitate) cannot be important consequences; on the contrary they are redundant or superfluous in a good sense, i.e. in the sense that something (proposition or predicate) which can be replaced by any arbitrary proposition or predicate – even by its own negation – without changing the validity cannot be important or essential in the inference (proof). In this sense a criterion which eliminates replaceable parts eliminates redundancies in the consequence class thereby preserving the informative consequence elements. Such a criterion can be formulated as follows:

R: α is an R-consequence of A iff both $A \mid\!- \alpha$ and it is not the case that a propositional variable (a predicate, the identity sign) is replaceable in α on some of its occurrences by any other propositional variable (predicate of same arity, two place relation) salva validitate of $A \mid\!- \alpha$.

That P2 (reducible parts) cause difficulties[25] can be seen from the fact that reversing the direction of reduction allows to produce disjunctions, to fuse separated parts (propositions with their predicates) into conjunctions (this again claims commensurability) and allows the problematic implications of the distributive laws. For the propositional part of PL1 it can be shown[26] that the restricted consequence class obtained by eliminating replaceable and reducible parts is PL1-equivalent to the full consequence class. This shows that nothing is lost of classical logic, only redundancies and superfluous elements in the consequence class are dropped out.

As to the question what remains of basic principles of logic – despite reasonable restrictions for applications in different areas – the following answer seems to be well justified: Applying Logic means applying at least the following principles:

[23] Many such difficulties have been shown by Weingartner, Schurz (1986, PSS); (1988, RCC); Weingartner (2000, RFC); (2001, ALO) in many different other areas like that of confirmation, explanation, law statements, disposition predicates, epistemic logic and deontic logic.

[24] cf. Mittelstaedt (1978, QLg), Chap. 2. (1998, IQM) eq. 4.52, Weingartner (2004, RSL) p. 245 and Chap. 13.1.2.2 eq(5*).

[25] See the above references and Schurz, Weingartner (1987, VDR) where it is shown that the problem of verisimilitude (in Popper's sense) can be solved by eliminating replaceable and reducible parts.

[26] The proof is given in Schurz, Weingartner (1987, VDR).

(1) The principle of logical consequence or logical deduction
An inference (deduction) is logically valid if it leads always from true premises to true conclusions. An inference is valid if there is no instance with true premises and a false conclusion. Moreover a valid inference permits to conclude from a false conclusion that at least one of the premises must be false. This principle can also be expressed model theoretically by saying that all the models which satisfy the premises (make the premises true) satisfy also the conclusion (make the conclusion true).[27]
(2) A tolerant version of the principle of non-contradiction: Two propositions of which one is the negation of the other cannot both be true (or more generally: cannot both have designated values). Observe that this principle does not presuppose bivalence and does not rule out many-valued logic (see Sect. 3.3.3).
(3) Some traditional rules of inference or logical laws which have a very simple and transparent structure like: dictum de omni,[28] modus ponens,[29] modus tollens,[30] hypothetical syllogism[31] (= transitivity of the implication), simplification,[32] identity laws.
This list cannot be extended arbitrarily since for instance bivalence or double negation is not uncontroversial (especially in intuitionism) and addition and adjunction make trouble in the application to the empirical sciences.

A modest step beyond these three principles (1)–(3) which is based on CL concerning the concept of validity, though solves most of the difficulties (in applications outside logic and mathematics) and deviates least from CL (having a consequence-class which is classically equivalent to the CL-consequences) seems to be restricted classical logic.[33]

What the above restrictions show again w.r.t. our original question is not that some laws of logic are refuted by empirical facts, but that not all laws of CL can be applied or used in a special field of application in the empirical sciences though they may be used in the application of formal sciences without any problems.

3.3.3 The Proper Domain

The proper domain of logic was described in the tradition (middle ages, already by Boethius) as that of concept, proposition and inference. And this

[27] The model-theoretic version was originally proposed by Tarski in 1936. Engl. Translation in Tarski (1956, CLC).
[28] $\forall x Fx \mid\!- Fa$
[29] $p \rightarrow q,\ p \mid\!- q$
[30] $p \rightarrow q,\ \neg q \mid\!- \neg p$
[31] $p \rightarrow q,\ q \rightarrow r \mid\!- p \rightarrow r$
[32] $p \wedge q \mid\!- p$
[33] For more on these problems see Sect. 13.2 on Quantum Logic.

is still today not a bad characterisation. It was also pointed out by Boethius and then by Avicenna that logic is concerned with the "second intentions": that is not with the concept of "blue" or the concept of "matter" but with the concept as concept or with the concept of concept. Similarly logic is concerned not with a particular proposition or inference but with proposition as such and its formal properties like true and false and its parts like subject and predicate and that one is predicated from the other; again not with a particular inference but with inference as such and its formal properties like validity and with the relation of the premises to the conclusion. The doctrine of the "second intentions" points to the high degree of abstraction of those entities with which logic is concerned: they are not concrete entities, they are not individuals on type level 0, but also not particular predicates on type level 1 or 2. A similar point is made by Russell concerning the entities of mathematics:

> "To begin with, we do not, in this subject, deal with particular things or particular properties... We are prepared to say that one and one are two, but not that Socrates and Plato are two... A world in which there were no such individuals would still be a world in which one and one are two."[34]

The high degree of abstraction and generality of the entities belonging to the proper domain of logic can be substantiated further by the following three principles E, M and L.

E: All concepts of logic are more general than all concepts of empirical sciences. All laws of logic are more universal than all laws of the empirical sciences.

This principle can be confirmed as follows: Every empirical concept like that of mass, energy, simultaneity, field, cell, metabolism, society, peace, etc. presupposes material carriers (bearers) or material processes, i.e. is matter dependent and can be defined only with material entities in spacetime. No concept of logic on the other hand is space time or matter dependent. All laws of the empirical sciences and laws of nature exclude at least some consistent and possible empirical structures (of an alternative world). But no basic law of logic like the principle of non-contradiction (in a tolerant form) or the principle of logical consequence excludes some possible (consistent) empirical structure.

Therefore it is clear that laws of logic cannot be laws of nature.

Concerning the application of the principle of non-contradiction to the empirical sciences one should distinguish different versions of the principle of non-contradiction (NC). They differ in strength such that some rule out many valued logics other do not.

NC1 $\neg(p \wedge \neg p)$, where p can only take the values true or false, is logically true;

[34] Russell (1919, IMP), p. 196.

NC2 $\neg(p \wedge \neg p)$ is logically true;
NC3 At least one member of the pair p, $\neg p$ must be false;
NC4 At most one member of the pair p, $\neg p$ can be true (or can have a designated value);
NC5 A proposition cannot be both true and false;
NC6 A proposition can assume only a single truth value.[35]

NC1 includes bivalence in the additional condition. Such a principle Leibniz seemed to have in mind as the principle of contradiction, since he gives two consequences of it: one is NC5, the other a clear formulation of bivalence or "tertium non datur" in his terminology.[36]

NC2 does not imply bivalence and holds in some systems of many valued logics, though not in all (it fails in the three-valued systems of Lukasiewicz, Post, Bochvar and Kleene).

Observe however that there are two different sorts of many valued logics: In the first sense "many valued" can mean: some values in addition to the values true and false. In the second sense "many valued" can mean more than one value of true and more than one of false such that there is nothing between true and false only different values for true and different ones for false.[37] In the latter case bivalence does not rule out many-valuedness.

NC3 is very restrictive as it requires values false in all lines of a matrix (truth table) in which the negation of the proposition has other values than false (for instance "indifferent").

NC4 is certainly the most tolerant version which was also defended already by Aristotle.[38]

It is with respect to this version that we can definitely say that it does not exclude any empirical structure.

NC5 is mentioned by Leibniz and is acceptable in general at least for all applications of logic to empirical sciences.[39]

NC6 finally expresses the view that when assigning truth values to propositions, one presupposes bivalence, i.e. the proposition, if it has value v cannot have another value v' $(\neq v)$.

[35] NC2–NC6 are mentioned by Rescher. He distinguishes another version of NC2 which differs from it only if the negation operator does not have the reflection property. cf. Rescher (1969, MVL), p. 144ff.

[36] cf. Leibniz (NE), 4, 2, 1. For a detailed discussion see Weingartner (1983, IMS), p. 160ff.

[37] Matrix systems of this sort have been used for independence proofs, for example by Bernays. A six-valued matrix system (three values for true and three for false) for a modal logic with 14 modalities has been proposed in Weingartner (1968, MLT).

[38] cf. Aristotle (Met) 1011b14 and 1062a22.

[39] Rescher describes some quasi-truth functional systems for which it does not hold. Ibid. p. 166ff.

To sum up it seems that for a logical system applied to natural sciences the principle of non-contradiction has to be asserted in the form NC4 whereas NC5 and NC6 can be accepted in the metatheory for the formation of the system.

We want to point out however that we do not think that nature or reality (or their real processes) violate NC4; violations of NC4 can only happen in human thinking and theorising about nature. Therefore paraconsistent logics, which allow a violation of NC4, can be understood as models of human thinking but not of real processes.

M: Some concepts of mathematics are less abstract than all concepts of logic. Some laws (or statements) of mathematics are less general than all laws of logic. The first part can be proved by the fact that the concept of a particular natural number, say 5, is less abstract than any concept of logic. The only concepts of entities obeying the uniqueness condition (which is satisfied by the natural numbers) in logic are the truth values T (true) and F (false) but logic can be built up dispensing with them, i.e. in a proof theoretical way. But for mathematics the entities of natural numbers are basic and indispensible. And moreover there have to be infinitely many in order to do mathematics. Concerning laws or statements of mathematics: Statements like $5 + 7 = 12$ of arithmetic, i.e. statements containing exclusively concepts of particular (single) entities – in this case natural numbers – (besides $+$, $=$) are genuine statements of mathematics. Examples of geometry show that there are statements which contain at least one concept of a particular kind or one constant like in : the sum of angles in a triangle equals 180°.[40] But no statement (or law) of logic contains concepts of particular entities. All genuine statements of logic contain variables (besides the logical terms like $\rightarrow$, $\wedge$, $\vee$, $\neg$).

L: Some concepts of logic are more abstract than all concepts of mathematics and some laws of logic are more general or more basic than all laws of mathematics. The first part can be shown to be true by taking the concepts of identity or predication (where a two or more place predicate is called a relation). The concept of identity is presupposed in mathematics and so is the general concept of predication (relation). Every mathematical equation is some more specific identity and every mathematical function is some kind of predication or relation. The second part is evident from the fact that the law of identity $x = x$ or the principle of logical consequence are more universal than any law of mathematics: the first is obeyed by every mathematical equation and the second by every mathematical proof.

For a general demarcation between logic and mathematics we propose the following conditions L1 to L4 for laws of logic and M1 to M3 for laws (theorems) of mathematics:

[40] For more see Sects. 4.1 and 4.2.

L1 Let s be a wff (well formed formula) in prenex normal form (i.e. all quantifiers – if any – at the beginning of the sentence). Then all expressions, except logical constants, are either universally quantified or can be universally quantified salva veritate (validitate) of s. Logical constants are: $\wedge$, $\vee$, $\rightarrow$, $\neg$, $\leftrightarrow$, $\in, =, \subseteq$, $\subset$, $\cap$, $\cup$, etc.

L2 The universe of discourse (the domain to which the variables refer) contains at least one object (i.e. is not empty). This holds for PL1 and all systems discussed in Sect. 3.32(3) above. A logic which permits an empty domain (depending on the inference rules or non-refering terms) an empty logic or a free logic. These systems have however some other properties.

L3 s is a valid inference rule which can be reinterpreted as a true universal implication which obeys L1.

L4 s is a logical consequence of a sentence obeying L1 or L3.

M1 Let m be a wff in prenex normal form and let all expressions of m (except the logical and mathematical constants like above and $<, >, +, -, \cdot, \emptyset$, etc.) have an index of stratification $i (0 \leq i < n)$ (0 for those of lowest type). Then there is at least one expression (usually predicate variable or set variable) which is existentially quantified and has index $i \geq 1$. Example: All axioms of standard set theory – except the axiom of extensionality – have such an existentially quantified variable representing the existence of the respective set (of pairing, sum set, power set, etc.).

M2 The universe of discourse contains at least denumerably infinite many objects.

M3 m is a logical consequence of a theorem of mathematics according to M1 and M2.[41]

3.3.4 No Laws of Logic are Laws of Nature and no Laws of Nature are Laws of Logic

This answer to the question "Are the laws of logic laws of nature?" can be substantiated as follows. The names "laws of logic" and "laws of nature" are taken from the proper domain of logic and from the proper domain of the natural sciences. But they are not taken from their domains of application. Thus although the laws of any science A which is applied to another one B contains variables the instances of which may be terms (signs) of entities belonging to a specific (proper) domain of a particular science B, such a law is not a law of that particular science B. Thus although the variables of $x = x$ may be interpreted as signs for physical entities either individual ones like

[41] For the demarcation in general see Quine (1970, PLg) and Russell (1919, IMP), p. 202ff. For free logics and empty logics see Bencivenga (1986, FLg). For the concept of stratification see Quine (1969, MLg), p. 157. For details on this demarcation and possible others see Weingartner (1982, DLM). For the axioms of standard set theories see Fraenkel, Bar Hillel, Levy (1973, FST) and Gödel (1940, CCH).

elementary particles or abstract ones like masses or fields, and although the physical entities obey this law, nevertheless this law can never be a law of physics. One important reason is that no law of logic excludes any consistent empirical structure but every law of physics (and of nature) does.

On the other hand no law of physics (and of nature) can be a law of logic. This is so because no law of physics can be universalised in such a way that by replacing physical terms by variables we get a law of logic. This just means that no physical law is a substitution instance of a law of logic. Thus a law of physics like $F = m \cdot a$ (Newton's second law)[42] although of course obeying the identity law $x = x$ and its instance $F = F$ cannot result in a law of logic just by replacing the three physically interpreted signs by three different variables. The same holds for other laws of natural sciences like those of chemistry and biology.

An analogous consideration can be made concerning mathematics: a certain homogenous partial differential equation of second order – as long as its terms do not have a physical interpretation – is not a physical law; even if physical entities will obey the equation if it is a mathematically valid equation. On the other hand a replacement of the physical terms in the Schrödinger equation by mathematical variables will not lead to a mathematical law; less than that would anybody count stationary states of energy to entities of the proper domain of mathematics. Though they may count as entities of the domain of application of mathematics. But as it was said above the name "law of logic" and also "law of mathematics" are not taken from their domains of application but from their proper domains.

From the above considerations it is plain that the proper domain of logic can clearly be distinguished from the proper domain of natural sciences. Laws of logic and laws of nature have no common proper domain. But although this is unambiguous w.r.t. the theoretical demarcation it can be hidden in the practical application. Thus it could happen that a special law of physics is not really a law of nature but merely a logical law disguised as a law of physics. In several fields of physics we can find disguised logical structures formulated in terms of physics. For example, the algebraic structures of the state spaces in classical and quantum mechanics are nothing else than Lindenbaum–Tarski algebras of calculi of the underlying propositional logic formulated in terms of classical or quantum mechanics. These structures are not proper laws of nature and they cannot be falsified by experimental evidence within their domain of application.

3.4 Answer to the Objections

3.4.1 (to 3.1.1) It is correct that the laws of logic are applicable to all objects dealt with by the natural sciences and consequently to all objects which are

[42] This law is written here in a rough form only but for the point to be made the form is irrelevant.

described and ruled by laws of nature; and moreover that these objects obey the laws of logic. But from this it does not follow that the laws of logic are also laws of nature. Since – as it was said in the answer – the names “law of logic” and “law of nature” are not taken from the domain of application; since then not only the laws of logic would be laws of nature but also the laws of nature would be also laws of ethics because human actions obey also the laws of nature. Therefore these names “laws of logic” and “laws of nature” are taken from the respective proper domains of the disciplines which are different.

The citation of Kant does not make the wrong claim that the laws of logic are laws of nature. It says only that the laws of logic are applicable to all objects whatsoever. However there is a problem with Kant’s view of laws in general: Whether he thinks that the a priori laws of logic are just laws of the a priori structures of our thought; since in an analogous way he claims that the “laws of nature” are just the a priori laws (structures) of the human mind who projects them on nature.[43] In any case we do not share this kind of subjective or idealistic view about laws of nature which is clear already from Chap. 2. Concerning laws of logic Kant mentions only the principle of non-contradiction, to which – according to him – all other laws of logic can be reduced and stresses that these laws are all analytic. The first (reduction) is of course not correct proof theoretically, and semantically there is no selection of a principle by a mark of distinction anyway. The second (analytic) is understandable, since Kant understood logic as syllogistics (and had apparently no knowledge of further developments in the Stoics, the Scholastics and in Leibniz). But every syllogistic mode has to obey the principle that the predicates in the conclusion (subject term and predicate term) are included in the predicates of the premises (subject term, predicate term, middle term). And this was just Kant’s concept of taking out (unfolding, extricating) the conclusion form the premises with which he described analyticity.[44] Applied to propositional logic the “analyticity condition” from syllogistics requires that there are no propositional variables in the conclusion (in the consequent of a valid implication) which are not already in the premises (antecedent). This is an interesting relevance criterion which was called Aristotelean criterion of relevance and has been investigated elsewhere.[45] For the quantificational part of PL1 however an analogous requirement of analyticity is not suitable, since it would generally forbid to introduce new individual terms in the conclusion; i.e. existential generalisations of the form $a = a \vdash \exists x(x = a)$ and many similar ones in the logic of relations would be forbidden.

---

[43] cf. Kant (1783, PzM), Sect. 36. This is however not the place to discuss this intricate question of the history of philosophy in detail.

[44] For the concept of analyticity in general and for an interpretation of Kant’s analyticity with the help of formal logic see Hintikka (1966, AAn), (1966, KVd).

[45] cf. Parry (1933, AAI), Weingartner (1985, SRC), which provides a semantics with matrices for this Aristotelean relevance criterion, and Weingartner, Schurz (1986, PSS).

3.4.2 (to 3.1.2) Although logic is the most general of all sciences with respect to application and although the laws of logic rule the elements of the universe of discourse of the natural sciences it does not follow that the laws of logic are laws of nature.[46] One important reason was given in the answer and in the commentary to the first objection: the domain of application is not identical with the proper domain of which the names "law of logic" and "law of nature" are taken. Another reason is that laws of nature satisfy all the conditions for genuine laws (see Chap. 2) but laws of logic do not satisfy all of them.

Still another reason can be grasped from what has been said in the section "Domain of Application" (3.3.2): Not all laws of logic are equally applicable to all fields of research. Whereas all laws of CL seem to be applicable in classical mathematics and set theory, not all laws of CL are applicable in all empirical sciences. And moreover some presuppositions of CL are not laws of logic, but strong assumptions for the formation of the system (calculus) like the one that every proposition (predicate) can be combined with every other.

The phase "the laws of logic rule the elements of the universe of discourse of natural sciences" should not be misunderstood. To "rule or order certain things" can have at least two different meansings: (i) to rule something by causally influencing it; (ii) to rule or order in the sense that the objects ruled satisfy the rule. Here it is only understood in the second sense (ii). That means nothing else than that the natural objects (for example material bodies) satisfy the laws of logic, i.e. are consistent objects (do not have incompatible properties at the same time) or satisfy the logical consequences if they satisfy the premises, etc. When it is said that laws of nature rule the elements of the universe of discourse of natural sicences then "law of nature" is understood in the sense of L2 (law like structure of nature) or in the sense of L4 (the true law) (cf. Sect. 1.3) but not in the sense of L3 (law statement) which is only an approximation to L4.

3.4.3 (to 3.1.3) Concerning this objection we have to distinguish two things: First the opinion that the laws of logic can be justified or revised by contingent facts (like perception and others). Secondly whether independently of such an extreme view of logic one can show that the laws of logic are not laws of nature. Concerning the first one can show that Russell's opinion leads to absurd consequences: Although one can admit that finitistically verifiable propositions concerning (the finite part of) natural numbers (i.e. the "multiplication table") play some role of direct evidence ("sense perception")

[46] Concerning logic this holds for PL1 (with identity). Concerning higher order, set theory in the version of Zermelo Fraenkel or Neumann Bernays Gödel seems to be the most general formal system. Maximal generality may hold also for an ontology or metaphysics which is understood in a very general way: everything or almost everything can be understood as some being (in the widest sense). Though also here important restrictions for consistency have to be made like those put on highly general classes, like the restrictions on the universal sets in set theory.

for elementary understanding[47] or even for building up the system of natural numbers in a constructivist way (cf. Chap. 4) it seems clearly untenable to extend such a "justification" to other domains of mathematics.[48] It can also not be acceptable for the generality of logical laws. Thus the logical law that all logical consequences from true premises are true cannot be justified by sense perception. Concerning the opinion of Quine one can agree to have a weaker system of logic (without bivalence) – i.e. to deviate from classical logic – for the application to certain domains outside logic, say mathematics from the intuitionistic point of view. But from this it does not follow that one can "revise" every law of logic. For instance in this sense one cannot "revise" the above principle of logical consequence or a very tolerant version of the principle of non-contradiction (recall Sect. 3.32 above). On the other hand one can even admit what is a common feature of all paraconsistent logics: to reject (certainly to reject as relevant, not necessarily to reject as valid) the inference that from a contradiction any statement whatsoever (not in any relation to the premises) can be derived which is a theorem of classical logic. According to the above restriction (see 3.32(b)) such a conclusion is a replaceable part. This move was criticised even by such a classical logician as Tarski:

> "A theory becomes untenable if we succeed in deriving from it two contradictory sentences. [...] People who are acquainted with modern logic are inclined to answer this question in the following way: A well-known logical law shows that a theory which enables us to derive two contradictory sentences enables us also to derive every sentence. [...]
>
> I have some doubts whether this answer contains an adequate analysis of the situation. I think that people who do not know modern logic are as little inclined to accept an inconsistent theory [...]; and probably this applies even to those who regard [...] the logical law on which the argument is based as a highly controversial issue, and almost as a paradox."[49]

Concerning the second point it can be said that even independently of the opinions of Russell and Quine laws of logic do not satisfy the necessary conditions for laws of nature and therefore cannot be laws of nature. In Chap. 2 it has been shown that laws of nature satisfy all the eight conditions which have been described for genuine laws. Now laws of logic w.r.t. their usual understanding do not satisfy conditions G2, G6, G7, G8: they do not describe conserved and invariant properties because they allow every property (do not exclude any except the inconsistent ones) (G2); they do not have empirical informative content (G6); they do not belong to a system of laws which makes

[47] cf. Feferman (1964, SPA), p. 3f.

[48] cf. the detailed critical remarks concerning an exaggeration for an analogy between mathematics and physics by Kreisel in his (1965, MLg), Chaps. 3 and 5, (1967, MLW), Chap. 2, and (1974, NMT).

[49] Tarski (1944, SCT), p. 367.

up the central part of an (empirical) theory (G7); they do not refer to mind independent objective reality in the empirical–ontological sense but to objective conceptual entities (G8).

3.4.4 to (3.1.4) The answer to this objection is clear from the section Domain of Application (3.32). The laws of logic are not laws of nature. First because the former are complete, the latter are not (subsection (2)). Secondly, limitations concerning the application (as described in subsection (3)(i) and (ii)) do not invalidate or limit the laws of logic themselves but limit their application and may restrict some presuppositions which are not laws of logic but formation rules for the propositions and predicates used in the laws. But the fact that those laws of logic which are not applicable in certain areas of empirical sciences are applicable without any problems or necessary restrictions in other areas (especially those of formal sciences) shows that they are not invalidated by the limitation of applicability: Contingent constraints of empirical facts lead to a selection of logical laws for the special application rather than to a violation. And since the so selected laws of logic are still belonging to the proper domain of logic (are still laws of logic) by other criteria they are not reducible to laws of nature.

3.4.5 (to 3.1.5) The answer to this objection and especially to the view expressed in the introduction and in later passages of Boole's *Laws of Thought* is similar to that given in 3.4.3 to Russell: the generality of the laws of logic cannot be ultimately justified by the operations of the human mind in the sense of subjective human experience, even in the most evident experience of these operations. Since – as it is well known from the history of mathematics and the sciences – some propositions which have been proposed as most evident axioms turned out to be inconsistent or incompatible with experimental results. On the other hand it can be conceded that evident operations of the mind play a role for the process of learning and understanding laws of logic.

That the laws of logic are independent from laws of nature or laws of thought (or operations of the mind) in an important sense can also be seen from an argument brought forward by Haldane which is directed against materialism: "if materialism is true, it seems to me that we cannot know that it is true. If my opinions are the result of the chemical processes going on in my brain, they are determined by the laws of chemistry, not of logic."[50] One may first object that also computers work in full accordance with the laws of logic. But to this objection there is a simple reply: The computer is programmed and constructed by humans in such a way that it uses the laws of CL (usually with some special restrictions like resolution rules, Horn clauses, etc. in order to avoid redundant inferences). One could construct a computer which makes invalid inferences or which makes valid inferences only at random. This shows unambiguously that the computer does not possess the logical laws in

[50] Haldane (1937, InM), p. 157. Popper discusses Haldane's argument and gives also reasons for the need of standards of validity. cf. Popper, Eccles (1985, SBr), p. 76f.

its nature. And moreover since both computers and humans (logicians, mathematicians and scientists) can and do commit logical (and other) errors, it is plain that by the laws of logic we understand standards of validity which have some important independence of laws of nature and of the physical world.

3.4.6 (to 3.2.1 and 3.2.2) The argument in 3.2.1 agrees with what has been said in our proposed answer. To the citation of Wittgenstein (3.2.2) we want to comment two things: First Wittgenstein uses the terminology he and people of the Vienna Circle invented and which was used afterwards by many logicians. Questions of terminology are not important except the terminology is misleading which is certainly the case with the word "tautology". Secondly to say "therefore the propositions of logic say nothing" is hardly acceptable as a conclusion from an invented terminology. Moreover the claim is a bold exaggeration or understatement. Since Frege and Russell have given logic a precise form a great number of interesting theorems has been proved (inside the system of PL1 and about the system). In a similar way proposition 6.2 of the *Tractatus* ("Mathematics is a logical method. The propositions of mathematics are equations, that is "Scheinsätze") is at least misleading if not plainly false. It seems to presuppose a strong "Logicism" (reduction of mathematics to logic) and seems to assume (wrongly) also for mathematics completeness and recursive enumerability of all true propositions as it is claimed for logic in 6.125 (*Tractatus*).[51] We shall not go into the question of whether the laws of logic are analytic – though widely and controversially discussed in philosophy. It suffices to say that as long as the claim of the analyticity of logic says only that the laws of logic do not describe and explain empirical (physical, biological…etc.) facts it can be easily agreed upon.[52]

[51] This fits to Wittgenstein's mechanical picture of logic, he had at the time of writing the *Tractatus*. Though correct for classical propositional logic the mechanical picture does not anymore fit to PL1 and much less to higher order logic because of its undecidability (though completeness).

[52] For a proposal to clarify different conceptions of analyticity by precise logical means see Hintikka (1966, AAn)

4

Are the Laws of Mathematics Laws of Nature?

In every law of physics mathematics is contained as an essential ingredient. If a physical law is a valid law of nature, then the mathematical laws contained in it hold a fortiori, indirectly, in the reality too. However, this very general aspect is not the problem that we are interested here. Instead, we will ask whether the mathematical laws which refer to the objective reality directly and without reference to physics, are valid laws of nature. Hence, we restrict our considerations to those parts of mathematics which are not concerned with abstract entities like groups, rings, and lattices but with concrete mathematical objects like natural numbers, geometrical objects in a three dimensional space and probabilities of real events. Here, we will investigate the question mentioned with respect to arithmetic of natural numbers, three-dimensional geometry and elementary probability.

Generally, we must distinguish two subquestions. Firstly, we ask, whether a given mathematical law holds in nature at all and secondly, whether this law, provided it holds in nature, is a genuine law of nature. Consequently, we proceed in two steps. In the first step we investigate the validity of the mathematical law considered in the objective reality and ask why this law holds in nature. On the basis of an answer to this question in a second step we will investigate whether the law in question is a genuine law of nature. As mentioned above (Chap. 2) a genuine law of nature is valid, it refers to the objective reality and the reason for its validity are certain features of the reality. A genuine law of nature is a contingent law and can – in principle – be falsified.

The three mathematical fields which we will study here can be formulated by means of convenient axioms. However, these fields are not genuine axiomatic structures but fields that can also be obtained by means of an operational approach. Within the framework of this approach we can find firstly an answer to the question why the laws considered hold for mathematical objects and secondly why these laws hold also for real objects. For the three fields mentioned we will show that under well defined conditions the laws of arithmetic, geometry and probability hold strictly in the real world, partly

for a priori reasons and partly for contingent reasons as the genuine laws of nature.

4.1 Arithmetic

In our everyday experience and in science we constantly apply the laws of arithmetic to the exterior reality. *Prima facie*, there are no doubts that the laws of arithmetic hold in our experience. We are able to count sheep in a flock of sheep and small planets in our solar system and we are sure that the numbers obtained are subject to the laws of arithmetic. Hence, the question arises whether the laws of arithmetic are contingent and falsifiable empirical results. Are they laws of nature – or are the laws of arithmetic a priori valid and hold, for this reason, also in our experience?

This question has a long history which will not be reported here. We mention only briefly that Kant considered the laws of arithmetic as a priori valid since they follow from the most general preconditions of possible experience.[1] Since Kant did not provide an explicit proof of this statement, i.e. he did not deduce Peano-like axioms, his argument was exposed to the critique of the logical empiricism of the 20th century. In particular, Quine[2] argued in favour of a fallibilism of arithmetic, such that arithmetic as a whole could be falsified by empirical evidence. In contrast, Wittgenstein pointed out that modifications of arithmetic for empirical reasons would lead to serious difficulties since they would presumably invalidate the preconditions of that experience which led to the modification in question.[3] Wittgenstein did not elaborate this argument in detail and in particular, he did not formulate explicitly the arithmetic preconditions of experience.

4.1.1 Question: Is Arithmetic a priori Valid?

In order to demonstrate the a priori validity of arithmetic and the impossibility of its empirical falsification, we will follow a more formal way of reasoning which allows to formulate in detail the preconditions of experience which are expressed by the laws of arithmetic.[4] In particular, it must be shown that the way to obtain empirical results about the objective reality always presupposes some arithmetical structures. Here, we will restrict our considerations to that part of arithmetic which deals with natural numbers $1, 2, \ldots, n, \ldots$ and with propositions like the equation $2 \cdot 3 = 3 \cdot 2$, etc. Of course, arithmetic is a much wider field that incorporates negative integers, rational numbers, and also real numbers. However, the transcendental way of reasoning, which allows

[1] Kant I. (1787, KRV); Kant, I. (1929, CPR).
[2] Quine, W.v.O. (1961, LPV), p. 47.
[3] Wittgenstein, L. (1974, GdM), p. 97 and 161.
[4] Tetens, H. (1994, AAp).

to demonstrate – in the spirit of Kant – the a priori validity of arithmetic in the empirical reality, is restricted to the infinite set of natural numbers. Only within the limited framework of arithmetic of natural numbers we can hope to show the a priori validity of arithmetic and thus to answer the question whether the laws of arithmetic are genuine laws of nature.

In the sense of operational mathematics,[5] which we will adopt here, numbers are given by figures that can be constructed by the calculus of arithmetic. Figures which represent numbers, i.e. numerals, are either the usual ciphers 1, 2, 3,... or simply sequences |, ||, |||, $\cdots$ of dashes. A number system N of this kind is subject to the following rules.

N1 In any number system N there is a well defined beginning sign α.
N2 To any sign β there is a uniquely defined successor $\sigma(\beta)$, which is different from all the preceding signs.
N3 Any sign β of N can be constructed by repeated application of the successor rule N2 to the beginning sign α.

These rules are a formal description of the counting process. A number system N which fulfils the rules N1–N3 is an ordered system of signs which will be called "Peano sequence".[6] The elements of a Peano sequence are signs (numerals) of natural numbers and these numbers are subject to the well known Peano axioms of arithmetic. The Peano axioms read:

P1 0 is a natural number;
P2 to any natural number n there is a uniquely defined successor $\sigma(n)$ which is again a natural number;
P3 for any natural number n holds that $n \neq \sigma(n)$;
P4 for any two numbers n and m holds that $n \neq m$ implies $\sigma(n) \neq \sigma(m)$;
P5 if a proposition $A(x)$ (x is a variable for natural numbers) is fulfilled by $x = 0$ and if $A(x)$ is fulfilled by $x = \sigma(n)$ if it is fulfilled by $x = n$, then $A(x)$ is fulfilled by any natural number.[7]

However, there is a large ambiguity in the signs which denote natural numbers. We could use the usual ciphers, sequences of dashes, or material objects provided they can be incorporated into a Peano sequence. In any case, there must be a one-to-one correspondence between numbers and signs for numbers. Moreover, any empirical object which can clearly be distinguished from other objects, can be inserted into a Peano sequence and used as a sign for a number. Hence, a Peano sequences could – in principle – be constituted exclusively by means of distinguishable single objects, provided there are infinitely many objects available, what will never be the case. In other words, to say that to

[5] Mainzer, K. (1984, OpM), p. 806.
[6] Tetens, H. (1994, AAp).
[7] This axiomatic system differs from the rules N1–N3 by the "axiom of induction" P5, since within the operational approach to arithmetic axiom P5 is an obvious consequence and need not to be formulated explicitly.

any material numeral n there is – according to P2 – a successor $\sigma(n)$, is an idealisation.

4.1.2 Proposed Answer

The preceding arguments show in which way the laws of arithmetic of natural numbers can be justified. Propositions of arithmetic deal with natural numbers or, more precisely according to Hilbert[8] with signs for numbers. These signs are constructed by means of the rules N1–N3 and fulfil – for that reason – the Peano axioms P1–P5. Hence, numbers, which are denoted by numerals, ciphers etc. are subject to the laws of arithmetic. It is obvious that this way of reasoning, which demonstrates the a priori validity of arithmetic, does not depend on experience in any way. Hence, our first – but still preliminary – result is that *the laws of arithmetic are not laws of nature.*

4.1.3 Question: Is Arithmetic Valid in the Real World?

Next, we consider the problem of applicability of arithmetic to empirical objects. Why can the laws of arithmetic be applied to objects of reality? In order to answer this question we argue in the following way. Our experience of the exterior reality consists of individual objects which are carriers of observable properties. These objects can be distinguished from other objects and re-identified at a later time. Empirical objects of this kind appear in our experience as a temporally ordered sequence and this sequence has obviously the essential properties of a Peano sequence. However, real objects can change their properties and they could even be destroyed. Hence, they are not very useful as signs for numbers. The individual and distinguishable objects which we observe at certain time values can, however, be named by dual numbers, say, which contain all the information about the object including the time of its appearance. This ordered system of names has again the essential properties of a Peano sequence and it fulfils as any other Peano sequence the arithmetical laws, which follow from the Peano axioms.

Hence, if there are single objects in our experience, which can be distinguished and re-identified, then these objects can uniquely be named such that the names can be ordered as a Peano sequence. Since this Peano sequence fulfils the laws of arithmetic we conclude that the laws of arithmetic hold also for the named real objects. For this conclusion, we must only assume that any object, which appears in our experience can conveniently be named. However, this assumption seems to be trivial, since otherwise the object would not be subject of objective experience. Hence, we find that the laws of arithmetic hold without any restriction for objects of experience.

[8] Hilbert, D. (1922, NdM).

4.1.4 Final Answer

On the basis of this result we can try to find a final answer to the question whether the laws of arithmetic are genuine laws of nature. The laws of arithmetic of natural numbers hold for numerals on account of their construction and thus for individual objects named by these numerals. They are laws in the sense of L5. Hence, the laws of arithmetic of natural numbers are not genuine laws of nature, although they hold unrestrictedly for the single, distinguishable objects of our experience. Arithmetic of natural numbers is the most important and most simple example, which shows, that unrestricted validity of some laws in our experience does not imply that these laws are genuine laws of nature. It could happen, that they can be justified by a priori reasoning, as in case of arithmetic.

4.1.5 Reservation

There is, however, still a last reservation against this way of reasoning that does not directly invalidate our final result, but rather the limits of applicability of these arguments. The unrestricted validity of arithmetic in the real world follows, if we assume that any conceivable experience consists of individual and distinguishable objects. Usually, this presupposition is taken for granted.[9] If, however, in contrast to this assumption our experience were not composed of single and distinguishable objects, which can conveniently be named by numerals, then the whole way of reasoning presented here would be without object.

A situation of this kind seems to be realised in quantum physics, where experience consists of a set of objects (atoms, electrons, etc.) that are highly correlated and indistinguishable.[10] Hence, they cannot be named. Generally, quantum objects cannot be individualised and re-identified at a later time. There is, however, a remaining arithmetic component. The total number of objects is still a meaningful concept. Hence, for the constitution of this kind of residual experience presumably the cardinality of natural numbers is sufficient. Presently, it is not known how a reduced arithmetic in this sense would look like and in which sense it may be considered as a precondition of quantum physical experience.[11]

4.2 Geometry

In the history of physics and geometry there is a long lasting debate about the problem whether the laws of geometry are valid in the real three dimensional

[9] e.g. Tetens, H. (1994, AAp).
[10] Esfeld, M. (1999, HdQ).
[11] A "non-Peanoan arithmetic" of this kind could be called "quantum arithmetic".

empty space.[12] Whereas Kant considered the laws of geometry as a priori valid and applicable in the real world, the situation was changed essentially by the discovery of the non-Euclidean[13] and the Riemannian geometry.[14,15] On the basis of this discovery, Gauß as well as Riemann considered it a meaningful question which one of the many possible Riemannian geometries actually holds in the real space. However, this question cannot easily be answered experimentally, since the geometry of the real space is not directly observable. Of course, we can investigate by measurement the properties of a real triangle made of wood, iron, or of light rays. However, experiments and measurements of this kind will tell us only something about the geometrical properties of light rays and massive bodies, but nothing about the geometry of the empty space.

In this situation, we could go back to the preconditions of the geometry in space and investigate the problem whether these preconditions can be tested experimentally. We will not discuss here the most general geometry of the three dimensional space but restrict our considerations to the Riemannian geometry and to the three special cases of Riemannian geometry with constant curvature, the Euclidean, elliptic, and hyperbolic geometry. The reason for this restriction is that the geometries mentioned can conveniently be characterised by conditions which must be fulfilled by the measuring roads. These conditions, which seem to be very intuitive, follow from mathematical results by Helmholtz, Lie, and Weyl.

Helmholtz considered the distance between two points as the fundamental concept of geometry. In order to measure a real distance in the three dimensional space we must compare this distance with some unit length, i.e. we must move a measuring rod from the distance to be measured to the unit length. This procedure presupposes that the measuring rod is not changed by this transport in any way, i.e. that it is an ideal rigid body. Hence, a necessary precondition of measuring distances is the free mobility of rigid measuring rods in the real space without thereby changing their form in any way. It turns out that this precondition of free mobility has important consequences for the geometry of distances measured by roads.

---

[12] Mittelstaedt, P. (1989, PMP), p. 49 ff.

[13] Gauss, C.F. (1828, Dgc). Gauß meditated about the problem at least since 1800 but the early efforts (cf. a letter to W. Bolyai of 1804) were still directed to prove the truth of Euclid's fifth axiom. Later letters to Bessel (1829) and Schumacher (1831) show some ideas but without proofs. The first fully developed systems of "non-Euclidean geometry" are Lobachevsky (1829, OFG) – after having given a lecture about this topic at the section of Mathematics and Physics of Kazan university already in 1826 – and independently J. Bolyai's *The Absolute Geometry* of 1831, cf. Meschkowski (1978, PNM) p. 28 ff. and Bonola (1955, NEG) p. 84 ff.

[14] Riemann, B. (1854, HyG).

[15] Mainzer, K. (1980, GdG).

The condition of free mobility of rigid measuring rods was first expressed mathematically by Helmholtz[16] and reformulated more precisely by Lie[17] and Weyl.[18] Instead of a rigid body, we consider a set of points in a finite region of space. The free mobility of a body then corresponds to an isometric mapping of the region in space to itself. First, we assume that the region in space is sufficiently small, i.e. only arbitrarily small bodies are freely movable. Then we have the following:

Theorem I. (Helmholtz)[19]
If arbitrary small regions of space can be mapped isometrically to itself by a point rotation, then the space is Riemannian and the line element assumes the form $ds = (g_{\mu\nu} dx^{\mu} dx^{\nu})^{1/2}$.

If, however, the premise of this theorem is strengthened such that even finitely extended regions of space can be mapped isometrically to itself by a point rotation, then we arrive at

Theorem II. (Helmholtz)
If finitely extended regions of space can be mapped isometrically to itself by a point rotation, then the space is a Riemannian space with constant curvature, i.e. Euclidean $(K = 0)$*, elliptic* $(K > 0)$*, or hyperbolic* $(K < 0)$.

For the empirical geometry the two Helmholtz theorems lead to the following corollaries:

Corollary I. *If we are given sufficiently small, rigid bodies, freely movable in space, then the geometry of the space that is measured with these rigid bodies is Riemannian.*

Corollary II. *If we are given even rigid bodies of finite extension that are feely movable, then the geometry of the space measured with these bodies is the geometry of an Euclidean, elliptic or hyperbolic space.*

These results suggest that we are confronted here with the following situation: On the one hand, the geometry of the real space follows from the preconditions that distances can be measured at all, i.e. from the free mobility of the measuring rods. Hence if we are able to measure distances at all, then the laws of the corresponding Riemannian geometry hold a priori for the measured distances. On the other hand, the free mobility and rigidness of sufficiently small or finitely extended rigid bodies is an contingent and falsifiable empirical property which indicates that the laws of geometry are partly empirical laws of nature.

This is, however not entirely correct for the following reasons. An experimental test of the rigidness of a body can be performed only by marking

[16] von Helmholtz, H. (1868, TdG).
[17] Lie, M.S. (1888, TfG).
[18] Weyl, H. (1923, MAR).
[19] For the proof cf. Laugwitz (1960, DfG), p. 145 ff.

on the body investigated at least four points that do not lie in a plane and by comparing the distances between these points before and after a certain motion. If the distances were unchanged, then the system of points would have been transformed to another congruent system of points and the body could be considered as rigid. However, for demonstrating the congruence of two point systems one needs freely moving measuring rods the rigidity of which, in turn, could only be demonstrated by congruence measurements. Apparently, the rigidity of a measuring rod cannot be tested empirically in this way.

Under these circumstances, one can conversely assert of an arbitrary, freely movable body that it is rigid. In particular, according to Helmholtz, one could take a body that is rigid in the sense of physical laws. More precisely, this means that the body investigated must be rigid with respect to the totality of physical laws, which are, in turn, be formulated within the framework of that geometry which one obtains by using the body considered as measuring rod. Hence the test of the rigidness of the body will consist in demonstrating the consistency of the assumed rigidness, the corresponding geometry and the laws of physics. Obviously, the essential assumption which must be made here, is the existence of objects which fulfil this consistency requirement.[20] One can then use a body which is rigid in the sense of physics as measuring rod. If the body is sufficiently small, then the measured geometry of space is Riemannian; if furthermore it has finite extension, then one obtains the geometry of a Riemannian space with constant curvature. However, the geometrical propositions which one obtains in this way do not deal with the properties of empty space but rather with the properties of the most general measuring rods which are rigid in the sense of physics.

In conclusion we find that the laws of geometry in empty space depend on some transcendental arguments and on the presupposition that there are freely movable measuring rods which are rigid in the sense of physics. Hence, the laws of geometry in empty space hold partly a priori and partly for empirical reasons and are thus partly laws of nature.

4.3 Probability

Outline of the Problem

In many fields of physics, in thermodynamics, statistical mechanics, and quantum mechanics the experimental results do not consist of single events E_i but of probabilities $p(E_i)$ for the occurrence of events E_i, or more precisely of relative frequencies $p_i = f(E_i)$. These relative frequencies are usually treated by means of probability laws – provided the number n of events is sufficiently

[20] There are some doubts, whether objects of this kind which are defined exclusively in terms of classical concepts are still meaningful in the realm of quantum physics.

large. Hence, the laws of probabilities are in fact applicable to the real world. In addition, they fulfil also the other necessary requirements of law of nature. As in the preceding sections about arithmetic and geometry, we will investigate here the question whether the laws of probability are genuine laws of nature. The question turns out to be a rather intricate problem, which requires a more detailed investigation.

From Kant to Boole

In his *Treatise of Human Nature* (1739) David Hume emphasised that we never observe objects but only qualities and that it is nothing but imagination if we regard the observed qualities as properties of an object. Hence, any scientific cognition begins with the observation of qualities and it seems to be merely a question of interpretation whether in addition to the observed phenomena a fictitious object, “an unknown something”, is used for the description of experimental results. Obviously, there is no reason to expect that general laws like conservation of substance or causality hold in nature.

The same problem was treated by Kant in the *Critique of Pure Reason* (1787). However, in contrast to Hume, Kant emphasised that “objects of experience” are not arbitrary imaginations but entities that are constituted from the observed data by means of some well-defined conceptual prescriptions, the categories of substance and causality. Hence, the interpretation of the empirical data as properties of an object can only be justified if the object as carrier of properties is constituted by these categories. Kant formulated necessary and sufficient conditions that must be fulfilled by the observational data if the measurement results are to be considered as properties of an “object of experience”.[21]

This way of reasoning can be illustrated within the framework of classical mechanics. If we possess some objective cognition that relates to the external reality and not to the observing subject, then the observations in space and time must be connected by mechanical laws, which are specifications and realisations of the general laws of causality and of the conservation of substance. For example, in the planetary system there is a large number of small planets that can be observed only occasionally. Observations that can be obtained within a period of several months, say, consist of many isolated light points without any obvious connection. However, if these observed data refer to a well-defined astronomical object, then the light points must be points on a

[21] Kant did not claim that an interpretation of this kind is possible for any set of observations. However, “if each representation were completely foreign to every other, standing apart in isolation, no such thing as knowledge would ever arise. For knowledge is [essentially] a whole in which representations stand compared and connected”. In other words, without causal correlation the observed qualities cannot be considered as properities of an object of experience (Kant, I. (1929, CPR), A 97). This case of “empty knowledge” will become interesting also for the problem of probability attribution.

spacetime trajectory that is determined by Newton's equation of motion. This mechanical law is strictly causal and it conserves substance, i.e., it preserves the mass point as the carrier of the mechanical predicates.

The physical sciences such as classical mechanics, which are governed by strict causality laws, were extended in the nineteenth century by the incorporation of statistical theories. In these new fields, the primarily given entities are no longer events E_i in space and time whose causal connection can easily be recognised but probabilities $p(E_i)$ for the occurrence of events E_i or, more correctly, the relative frequencies of events E_i. Under these conditions the original Kantian way of reasoning is no longer applicable. If causal correlations between several events are not discernible, the objectivity of our cognition cannot be guaranteed in the Kantian way. Instead, one has to begin with the only perceptible empirical structure, the relative frequencies or probabilities $p(E_i)$ of various events E_i, and ask under which conditions probabilities of this kind refer to a certain exterior object. [22]

This problem was first studied by George Boole, best known as one of the founders of modern logic. In his book *The laws of Thought*[23] and in a subsequent paper *On the Theory of Probabilities*,[24] Boole investigated what he called "the conditions of possible experience". The problem treated by Boole can be sketched as follows. Assume that the observer is given a set of numbers p_1, $p_2, \ldots, p_n$ with $0 \leq p_i \leq 1$, which represent the relative frequencies of n merely *logically connected* events E_1, $E_2, \ldots, E_n$. One can then ask for the necessary and sufficient conditions that must be fulfilled if the numbers p_i are to be considered as probabilities for properties (given by E_i) of some object of experience.[25]

Classicality Conditions

In order to realise Boole's project, let us consider a sequence of n rational numbers $p_1, p_2, \ldots, p_n$, which correspond to the relative frequencies of the logically connected events $E_1, E_2, \ldots, E_n$. "Logically connected" means in this context that the events E_i are not logically independent but connected by logical operations and relations. The question that will be investigated here reads: what are the necessary and sufficient conditions that permit the interpretation of the numbers p_i as "probabilities" that can be attributed to some physical system? In other words, what are the conditions for the existence of a classical probability space such that there are n logically connected events

[22] We will restrict our considerations to the "relative frequency interpretation" of probabilities. "Single case interpretations" which make use of the concepts of "propensity" (Popper) and "potentia" (Heisenberg) might be of interest in fields different from natural sciences and will not be considered here. (cf. Mittelstaedt; P. (1997, QPT)).

[23] Boole, G. (1854, LoT).

[24] Boole, G. (1862, ToP).

[25] cf. also Pitowsky, I. (1994, CPE).

$E_1, E_2, \ldots, E_n$ whose probabilities $p(E_i)$ are given by the observed numbers p_i?

If the events E_i are logically independent, i.e. if they are not logically connected in any way, then the only conditions that must be fulfilled by the probabilities p_i are $0 \leq p_i \leq 1$. Clearly, these conditions are always fulfilled by the observed relative frequencies. A situation in which the observed data – the events E_i – do not show any obvious connection in either a causal or a logical sense can be compared with the "empty knowledge" case mentioned by Kant. Indeed, if each event were completely isolated from every other, "no such thing as knowledge could ever arise". In this case, the numbers $p_i (0 \leq p_i \leq 1)$ could not be attributed to an object system as the probabilities of its various properties.

If, however, the events E_i are logically interconnected, then one can investigate the question of whether the observed relative frequencies $p_i = f(E_i)$ of the events E_i can be attributed to an object system as probabilities for the properties given by E_i. For a more rigorous treatment of this problem, one has to specify first the logic of events and second the probability measure in question. Here we will give a brief account of the basic concepts needed in the following discussion. For more details and formal proofs we refer to the literature.[26]

An *event system* L is a triple $L = \langle L_0, \leq, \neg \rangle$ where L_0 is a set of (elementary) events that is equipped with a partial ordering relation $\leq$ and a one-place operation $\neg$, the complementation, such that L is an orthocomplemented orthomodular, partially ordered set which can be extended to a lattice by means of the two place operations $\wedge$ and $\vee$. In particular a system L is called a *classical event system* if it is a complemented distributive lattice with respect to the relation $\leq$ and the complementation $\neg$. Clearly, the logical meaning of the relation $\leq$ is the implication, the operation $\neg$ is the negation, and the operations $\wedge$ (and) and $\vee$ (or) can be defined as infimum and as supremum, respectively.

A *probability measure* p on the event system L is a map

$$p : L \rightarrow [0, 1]$$

of the set L of events onto the interval [0, 1] which satisfies the *Kolmogorov axioms*

K1 $p(\mathbf{0}) = 0, \qquad p(\mathbf{I}) = 1$
K2 $p(\neg a) = 1 - p(a) \quad$ for all $\quad a \in L$ (K)
K3 $p(a_1 \vee a_2 \vee \cdots \vee a_n) = p(a_1) + p(a_2) + \cdots + p(a_n)$ whenever $a_i \leq \neg a_j$ for $i \neq j$.

By an *event probability space* W we understand a pair $W = \langle L, p \rangle$, where L is an event system and p a probability measure on L. By means of the concepts L, p, and W we can now formulate the main problem.

[26] Beltrametti, E.G., Maczynski, M.J. (1991, CNP), Pitowsky, I. (1989, QPL).

A correlation sequence $K = \{p_1, p_2, \ldots; p_{ij}, \ldots\}$ is a set of elementary probabilities p_i and joint probabilities p_{ij}, where not all pairs (i,j) need to appear. A sequence K is called *consistently representable* if there is an event probability space $W = \langle L, p \rangle$ and a sequence of events $(a_1, a_2, \ldots, a_n)$ in L such that

$$p = p(a_i), \quad p_{ij} = p(a_i \wedge a_j)$$

whenever the pair (i,j) appears in K.

In particular, K is called classically representable if L is a classical event system. This means that the elements of K – the observed relative frequencies – may be considered as classical (Kolmogorovian) probabilities that refer to the elements a_i of a classical event system L_B. The property of a correlation sequence K of being consistently representable in a classical event space is called "*classicality*". The (necessary and sufficient) conditions under which a probability sequence K possesses the property of classicality are called "*classicality conditions*". The classicality conditions show that a given sequence of probabilities provides some objective knowledge about a physical system. Indeed, if we are given (by experiment) a sequence of probabilities $p_i = p(E_i)$ of events E_i, then the necessary and sufficient conditions for these events to indicate classical properties $P_i(S)$ of a physical system S are the classicality conditions.

In order to illustrate the concept of classicality we consider a probability sequence with two properties. For the properties, we write a and b and for the probabilities $p(a)$ and $p(b)$, respectively. For the joint probability (of the property $a \wedge b$) we write $p(a, b)$. The simplest example of a probability sequence is then given by (index "2" stands for two properties)

$$K_2 = \{p(a), p(b), p(a, b)\}.$$

Classicality Theorem T_2.

The correlation sequence K_2 is consistently representable in a classical event space if and only if the "classicality conditions"

$$\begin{aligned} &0 \leq p(a,b) \leq p(a) \leq 1 \\ &0 \leq p(a,b) \leq p(b) \leq 1 \\ &p(a) + p(b) - p(a,b) \leq 1 \end{aligned} \tag{C$_2$}$$

hold.

For the proof of theorem T_2 one needs the probability axioms (K) and some properties of the Boolean lattice $L_B^{(2)}$ that is generated by the two elementary propositions a and b. The two parts of the proof (that conditions (C_2) are necessary and sufficient)[27] illustrate the meaning of the classicality condition (C_2): If the probabilities of a correlation sequence $K_2 = \{p(a), p(b), p(a, b)\}$ pertain to a physical system as probabilities of its properties, then the properties fulfil the laws of classical logic and the probabilities

[27] Mittelstaedt, P. (1998, IQM), p. 130 ff.

are subject to the Kolmogorov axioms. From these premises the classicality conditions can easily be derived and vice versa. It should be mentioned that the whole way of reasoning is not restricted to two elementary propositions a and b, but can be extended to correlation sequences K_n, classicality conditions C_n and Boolean lattices $L_B^{(n)}$ that are generated by n elementary propositions. In particular, the case $n = 3$ plays an important role in quantum mechanics, where the conditions C_3 are usually called "Bell's inequalities".[28]

On the basis of these results we come back to our main question whether the probability laws are laws of nature. Within the framework of the present discussion probability laws are consequences of the Kolmogorov axioms, e.g. inequalities or equations. However, probability laws are not contingent empirical laws that could be falsified by experimental evidence. Instead, the empirical validity of probability laws is based on some kind of transcendental argument: If we are given a sequence $\sum_n = \{E_1, E_2, \cdots, E_n\}$ of events which refer to an object system S as its properties, then the relative frequencies $p_i = f(E_i)$ of events fulfil the laws of Kolmogorov probabilities.[29] However, we have also seen that not arbitrary sequences and their relative frequencies are subject to the probability laws. Moreover, correlation sequences K_n of probabilities which are given empirically, can be attributed to an object S if and only if the corresponding classicality conditions C_n are fulfilled. Hence the classicality conditions C_n are necessary and sufficient preconditions for the attribution of relative frequencies to an individual object. Consequently, these conditions C_n indicate the limits of validity and applicability of classical Kolmogorov probability. It is well known that some conditions C_n, in particular the Bell inequalities C_3 are violated in quantum mechanics.

Hence, probability laws are not genuine laws of nature that can be tested empirically. They hold whenever they can be applied. However the necessary preconditions under which probability laws can be applied to the physical reality, the classicality conditions C_n, are testable properties of the physical reality. They are fulfilled in classical physics and partly violated in quantum mechanics.

4.4 Concluding Answer

Are the laws of mathematics laws of nature? In the three cases discussed here we found similar but rather complex answers. On the one hand, the laws of arithmetic hold on account of the construction of its elements, the natural numbers, the laws of geometry hold for measured distances because of the free mobility of measuring rods and the probability laws hold since observed relative frequencies can be attributed to an individual object as probabilities.

---

[28] Mittelstaedt, P. l.c., p. 98 ff.

[29] To be more precise, the probability laws hold for sequences $\sum_n$ in the limit $n \to \infty$ except for non-random sequences.

Hence in the three cases the laws mentioned hold a priori as consequences of the preconditions of counting objects, measuring distances and attributing relative frequencies to a referent system. Consequently, under this aspect, they cannot be considered as contingent and falsifiable laws of nature.

On the other hand, the a priori validity of the laws considered depends on preconditions of possible experience that are by no means self-evident. The existence of individual objects, the existence of freely movable measuring rods which are rigid in the sense of physics, and the validity of classicality conditions for observed relative frequencies are testable features of the physical reality and even violated or at least questioned in some domains of this reality. Hence, arithmetic, geometry and probability considered as applied sciences depend also on contingent properties of the physical reality and are – under this aspect – partly laws of nature.

Part II

Properties of Laws

5

Does Every Law of Nature Express an Invariance (Symmetry)?

5.1 Introduction. Arguments Pro and Contra

Terminological Remarks

(1) The expression "invariance" and "invariant" will be used only for laws and for physical parameters and magnitudes (see also Chap. 6). However, the expressions "symmetry" and "symmetrical" are usually applied to both laws and physical systems. This holds also for their negations like "asymmetrical" and "not-symmetrical". The distinction between these two applications will be discussed in Sect. 5.3.1. If there is no danger for a confusion this two-fold usage is also adopted here.

(2) If the expressions "symmetry" and "symmetrical" are applied to laws then we use the expressions "invariance", "invariant" on the one hand and "symmetry", "symmetrical" on the other as having the same meaning. Under this condition "symmetry-breaking" will mean the same as "violation of invariance".

(3) The terms "symmetrical", "not-symmetrical", and "asymmetrical" are used according to the following definitions:
   (i) Relation (function) R is symmetrical if and only if $\forall x \forall y (Rxy \rightarrow Ryx)$
   (ii) Relation (function) R is not-symmetrical if and only if $\neg \forall x \forall y (Rxy \rightarrow Ryx)$
   (iii) Relation (function) R is asymmetrical if and only if $\forall x \forall y (Rxy \rightarrow \neg Ryx)$
   (iv) Relation (function) R is non-symmetrical if and only if R is neither symmetrical nor asymmetrical.

   From these definitions it follows:
   R is asymmetrical and R is not empty $\rightarrow$ R is not-symmetrical;
   R is non-symmetrical $\rightarrow$ R is not-symmetrical;
   R is not-symmetrical $\rightarrow$ R is asymmetrical or R is non-symmetrical.

Examples: x, y are physical systems and R is translation or rotation or the relation between right and left, past and future.

Arguments Contra

5.1.1 If every law of nature expresses an invariance (symmetry) then there must be a selected number of mutual independencies specific for that law among some selected variables of the law, and not just an arbitrary unlimited number of independencies. Thus for example in the hypothesis "the acceleration (a) of a freely falling body does not depend on its mass (m)" we mean that the values of a stay the same no matter what the values of m are. But, as Bunge pointed out, an unlimited number of mutual independencies could be set up concerning any given set of variables:

> "For instance, we might truly say that the acceleration of a freely falling body is independent of its colour, texture, price, aesthetic value and so on without end."[1]

This seems also to hold for laws of nature. Thus in the law "the total momentum (p) of a system of particles subject to non-frictional forces (f) is conserved" the values of p stay the same not only when the values of f change but also when an unlimited arbitrary number of other values of properties of the particles change.

Therefore: not every law of nature expresses an invariance (symmetry).

5.1.2 If asymmetrical states are realised in nature then laws describing them cannot be called symmetry principles or cannot express a symmetry. Now, as Bunge says, the following law is usually regarded as a law of nature: "In nature only symmetrical or antisymmetrical states are realized".[2]

Therefore: not all laws of nature can be called symmetry principles or do express a symmetry.

5.1.3 If a law of nature describes and explains asymmetric phenomena then this law cannot be called a symmetry (invariance) principle or does not express a symmetry. But several laws of nature describe and explain asymmetric phenomena. For example Newton's laws of planetary motion and also Kepler's describe and explain the elliptic orbits of the planets. These are however asymmetric solutions with respect to the most simple solutions according to which the orbits are circles due to rotational symmetry.

Therefore: Newton's and Kepler's laws of planetary motion do not express a symmetry or cannot be called symmetry (invariance) principles.

5.1.4 If all laws of nature express important symmetries then they have to be symmetrical under a mirror reflection (i.e. right-left symmetrical) which is

---

[1] Bunge (1967, SRI), p. 315. Though texture, price and aesthetic value are of course not properties of physical systems, the argument provokes the important question concerning the selection of independent parameters with respect to which laws are invariant.

[2] Bunge (1967, SRI), p. 368. What Bunge calls "antisymmetrical" is in our terminology "asymmetrical".

also called parity conservation. But parity conservation is violated as was discovered in 1956: theoretically predicted by Lee and Yang and experimentally confirmed by C. S. Wu and others in a cobalt-60 β-decay.

Therefore: not all laws of nature express important symmetries (invariances).

5.1.5 If all laws of nature express important symmetries then they have to be symmetrical under particle–antiparticle symmetry, i.e. they have to be charge symmetrical.

But the violation of charge symmetry was conjectured by Lee and Yang in 1957 and experimentally established by Christenson et al. and Bennett et al. in 1964 and 1967.

Therefore: not all laws of nature express important symmetries or invariances.

5.1.6 If the laws of nature express symmetries (invariance) then the demarcation between symmetry and symmetry-breaking must be based on something objective and not on assumptions about observability. But as Lee says this demarcation seems to be based on assumptions about what is non-observable and what is observable:

> "Since non-observables imply symmetry, any discovery of asymmetry (symmetry-breaking) must imply some observable".[3]

Thus the physical assumption that absolute position is a non-observable leads to three conservation laws or symmetry principles of momentum. And similarly with other such assumptions like direction in space, point of time, etc. Thus symmetry principles are dependent on assumptions about observability.

Since laws of nature cannot be dependent on assumptions about observability laws of nature cannot be called symmetry principles or cannot express symmetries.

5.1.7 If the laws of nature are principles expressing symmetry (invariance) then the demarcation between symmetry and symmetry-breaking must not be relative. But it seems to be an open question whether symmetry-breaking phenomena refute the respective symmetric (invariant) law or whether there is a symmetric law on the bottom and the symmetry-breaking phenomena are caused by special initial conditions with respect to which the law is invariant. For example the violation of parity or charge conjugation may be interpreted as refuting the respective symmetric laws or as caused by special initial conditions while keeping the underlying laws symmetric (invariant).

Since such interpretations seem to be relative with respect to the level of universality or abstraction the laws of nature cannot be principles expressing symmetry (invariance).

[3] Lee (1988, SAW), p. 11.

Arguments Pro

If in cases where a supposed law of physics turns out not to be invariant it should – according to the rules (norms) of methodology – be always replaced by a new law which specifies the rate of change then the laws of physics are invariance (symmetry) principles. As Popper says there should always be such a replacement:

> "Physicists as a rule hold that physical laws are eternal... It is indeed difficult to think otherwise, since what we call the laws of physics are the results of our search for invariants. Thus even if a supposed law of physics should turn out to be variable, so that (say) one of the apparently fundamental physical constants should turn out to change in time, we should try to replace it by a new invariant law that specifies the rate of change."[4]

Therefore: every law of physics is a principle of invariance (symmetry).

5.2 What a Law Is

Proposed Answer

Every law of nature expresses an invariance or symmetry. This can be seen as follows: Our understanding of any kind of genuine law is such that a genuine law is something which does not change, i.e. is invariant (symmetric) relative to something else which changes. But as it was shown in Chap. 2 every law of nature is a genuine law. Therefore every law of nature expresses an invariance.

Moreover it can be shown independently that every law of nature does not change relative to some parameters which change like space (position), time (point of time), inertial frame... etc. (see below). In this sense Wigner says:

> "It is not necessary to look deeper into the situation to realize that laws of nature could not exist without principles of invariance"

and

> "A law of nature can be accepted as valid only if the correlations which it postulates are consistent with the accepted invariance principles."[5]

Therefore every law of nature expresses an invariance or a symmetry. A more detailed justification of this argument is as follows: first a comment will be given to our understanding of a law which is connected with the philosophical tradition. Second different meanings of "invariance" and "symmetry" will be distinguished. Third it will be shown by discussing several groups of

[4] Popper, Eccles (1985, SBr), p. 14.
[5] Wigner (1967, SRf), p. 29 and 46.

symmetries (invariances) that laws of nature express an invariance or symmetry. Fourth the relation of symmetry and symmetry breaking will be discussed.

5.2.1 Our Understanding of What a Law Is

In addition to what has been said already in Chap. 2, G3, G4, the feature of invariance belongs to the oldest features for understanding what a law is. In order to be able to describe and explain movement we need to distinguish something which changes relative to something which does not change. This important distinction is pointed out by Aristotle[6] also as a criticism of Parmenides' theory of the universe which assumes only one being and nothing else.[7] That what changes, moves was thought to be contingent (not necessary) with respect to the not changing (or even not changeable) necessary principle or law. In general this idea belongs to the Greek Ideal of Science which was more or less manifest in several Greek thinkers from Thales on but was elaborated in detail by Plato and Aristotle: To describe and explain the visible (observable), concrete, particular, changing, material, contingent world by non-visible (non-observable) abstract, universal, non changing, immaterial and necessary principles.

In addition to this more general understanding the two particular invariances (symmetries) which are the oldest in the tradition are those of space and time. In some sense they are the most important invariance properties of laws of nature in general and of physical laws in particular.

> "The paradigm for symmetries of nature is of course the group of symmetries of space and time. These are symmetries that tell you that the laws of nature don't care about how you orient your laboratory, or where you locate your laboratory, or how you set your clocks or how fast your laboratory is moving."[8]

5.2.2 The Beginning of Time

The following passage points especially to the fact that the relations between the events which are described by the laws depend only on the intervals but not on a point of time when the first event occurred. That means that time-symmetric laws cannot designate or select a beginning in time or a first event:

> "Thus the time displacement invariance, properly formulated, reads: the correlations between events depend only on the time intervals between those events; they do not depend on the time when the first of them takes place.[9]

---

[6] Aristotle (Phys), 190a17f.
[7] Aristotle (Met), 986b15f. and Aristotle (Phys), 186a24ff.
[8] Weinberg (1987, TFL), p. 73. cf. Weingartner (1996, UWT) ch.5.
[9] Wigner (1967, SRf), p. 31.

The first philosopher who seems to have realised this very clearly was Thomas Aquinas. In his quarrel with Bonaventura at the university of Paris he defended the view that the beginning in time of the world (universe) cannot be proved from universal principles of, or laws about this world.[10] Because universal principles like real universal definitions and laws together with the universal concepts or predicates contained in them abstract from *hic et nunc* (here and now):

> "We hold by faith alone, and it cannot be proved by demonstration, that the world did not always exist. . . . . The reason is this: the world considered in itself offers no grounds for demonstrating that it was once all new. For the principle for demonstrating an object is its definition. Now the specific nature of each and every object abstracts from the here and now, which is why universals are described as being *everywhere and always*. Hence it cannot be demonstrated that man or the heavens or stone did not always exist."[11]

In this connection we want to mention that the question whether it can be demonstrated that the world has always existed or that it has a beginning in (with) time – answered differently by the two competing theories of the development of the universe – is a question about the completeness of the laws of nature – or at least of that laws we know. A system of laws L about a certain part P of reality is complete if and only if every truth about P is provable (derivable) from L.[12] Thomas Aquinas' stance was that the universal laws of nature (about this world) are not complete with respect to all questions (all truths) about this world. It is not just our insufficient knowledge of the laws of nature what he has in mind, but the true laws itself are incomplete according to him with respect to some special questions. That means that there are some statements about this world which are undecidable from the laws about this world. Or in more modern terms: the laws of nature are incomplete with respect to some important initial conditions.

[10] In contrast to Thomas Aquinas Bonaventura argued that an infinite past of the universe is *logically* impossible. This argument goes back to Johannes Philoponos *"De aeternitate mundi: contra Proclum"*. (1899, DAM).

[11] Thomas Aquinas (STh) I, 46, 2. Observe that by "demonstration" Thomas Aquinas means a rigorous proof from premises which essentially include necessary laws (of nature).

[12] For the question of the completeness of the laws of nature see Chap. 11. For the rôle of initial conditions see Chap. 8.

5.3 Invariance and Symmetry

5.3.1 Different Meanings of "Invariance" and "Symmetry" and their Context

We use the terms "invariance" and "symmetry" according to the terminological remarks stated at the beginning of this chapter: applied to laws the expressions "invariance", "invariant" and "symmetry", "symmetrical", respectively, are used as having the same meaning. For physical systems, however, only "symmetry" and "symmetrical" are used, but then with a different meaning.

In this section the distinction between different meanings of "invariance" and "symmetry" will be as follows: Concerning laws we shall distinguish three meanings of invariance (1–3):

Invariance (symmetry) in a very wide and general sense, invariance with respect to transformation groups and invariance of highest level principles in a hierarchical order. Further we shall make a remark on non-invariance (4). A further clarification will concern the relation between symmetric laws and physical systems on one hand and non-symmetrical phenomena on the other (5). Finally active and passive transformations will be related to the invariant laws (6).

(1) Invariance or symmetry in a wide sense
This type of "invariance" ("symmetry") is connected with our general understanding of what a law is: something which is stable, does not change, is independent relative to something else which is unstable, which changes, which is dependent. The rough idea is expressed above by describing the Greek ideal of science. But this rough idea can be made a bit more precise by the following consideration. We might ask the question: Under what kind of changes are laws of nature invariant? We might answer first by giving some of the continuous spacetime symmetries (see below 5.3.2). But then we might continue with discrete symmetries like CPT symmetry, etc. But where to end? This leads to the question: What is the set of all changes which do not change laws (of nature)?

Let us call the set of all the changes which do not change the laws (of nature) the symmetry group of nature. About this symmetry group Weinberg says:

> "It is increasingly clear that the symmetry group of nature is the deepest thing that we understand about nature today... Specifying the symmetry group of nature may be all we need to say about the physical world beyond the principles of quantum mechanics."[13]

[13] Weinberg (1987, TFL), p. 73. Weinberg's definition of "symmetry group of nature" is a little bit different and has some subjective element in it: "point of view" and "way you look at nature". "The set of all these changes in point of view is called the symmetry group of nature." cf. ibid. p. 72 and 73. We shall try to avoid this subjective element and use as a preliminary version: the set of all changes which do not change the laws. cf. Weingartner (1996, UWT) Chap. 7.

But how can we determine the set of all changes, which leave the laws invariant? This would mean to know the line of demarcation between contingent initial conditions and necessary and invariant laws. It would mean to know which constants when changed do not affect the laws and which do; and which initial conditions and boundary conditions would affect the laws when changed and which would not. Are the laws of nature invariant with respect to a change of the amount of energy (mass) of the whole universe (which is constant by the law of conservation of energy)? Or could we change the ratio of electron and proton mass slightly without changing laws?' From these questions it is clear that invariance (or symmetry) in this wide sense incorporates all the groups of symmetries which are listed in Sect. 5.3.2 below; and it may even concern further ones of which we are ignorant so far. However, this wide sense of invariance and symmetry should not be confused with a much more definite and restricted meaning to be described as follows:

(2) Invariance (symmetry) under changes of reference frames
This is the sense which is used nowadays to formulate most invariance (symmetry) principles (see (3) below) and to speak about the invariance (or symmetry) of laws: a law of nature is invariant (symmetric) if and only if it does not change under a change of (physical) reference frames. For example a law could be invariant under a change of inertial systems. In general a maximal set of transformations is combined to a transformation group, such as the Galilean group or the Lorentz group. It is said then that laws are Galilean invariant or Lorentz invariant, like those of classical mechanics and quantum mechanics on one hand, and those of classical electrodynamics, special relativity and general relativity on the other.[14]

(3) Invariance (symmetry) principles as laws about laws
Invariance (symmetry) principles are sometimes understood as meta-laws about laws (of nature) or as guiding principles which have to be satisfied by every genuine law of nature. Such a view is expressed in the following quotations of Wigner and Bunge:

"From a very abstract point of view, there is a great similarity between the relation of the laws of nature to the events on one hand, and the relation of symmetry principles to the laws of nature on the other.[15]

"... the new aspects which would be dealt with in these pages ... rather support ... and confirm the function of the invariance principles to provide a structure or coherence to the laws of nature just as the laws of nature provide a structure and coherence to the set of events.[16]

[14] For a definition of reference frame and coordinate system see Sect. 6.3.1 Invariance in the sense of (2) will be of special importance in Chap. 6.

[15] Wigner (1967, SRf), p. 16.

[16] Ibid., p. 17.

"It is good to emphasize at this point the fact that the laws of nature that is, the correlations between events, are the entities to which the symmetry laws apply, not the events themselves.[17]

"Apparently 'nonconservation of parity' refers to certain law formulas, and the asymmetry has testable consequences that can be compared with certain facts; in other words 'nonconservation of parity' is in this case an ambiguous phrase, since it refers both to certain laws and to certain sets of facts."[18]

Concerning this view of Wigner which he defends in more detail especially in Chap. 2 of the cited book we have to distinguish two things: (i) first the question of terminology and (ii) second the message of Wigner.

(i) Concerning terminology it is necessary to observe that there is a certain ambiguity in the application of the term "symmetry" (recall the terminological remarks at the beginning of this chapter). Symmetry and symmetry principles are also applied to real objects like snowflakes or letters and to physical systems like molecules or crystals. But here (in these quotations) they are used only as very abstract properties of laws of nature. This tells us that from the terminological point of view there is an ambiguity here which has to be taken care of and which can be made explicit (see Sect. (4) below).

(ii) Concerning the message of Wigner we have to distinguish two ways of understanding a law. First every law can be understood as an invariance principle, in the sense of expressing an invariance, along the lines which have been explained historically and systematically in Sects. 5.2.1 and 5.3.1(1) above. Secondly some laws can be understood as laws about laws or as meta-laws (or meta-nomological laws). And in this second sense Wigner understands "symmetry principle" or "invariance principle". Thus spacetime invariance can be expressed by the principle: All physical laws (or all laws of nature) are spacetime invariant. This meta-law can even be expressed as a normative principle in the sense of a valid rule for finding new laws or for requiring desirable properties of laws, or as a norm in the ideal sense: All laws or nature should (must, ought to) be spacetime invariant. And a law not satisfying this rule will then be thrown out of the laws of nature. Observe however that the prescriptive (normative) forms of such meta-laws are not testable, except in the sense that a norm cannot be valid if that what it requires is logically inconsistent or is not satisfiable by facts.

However it should be noticed that the point of Wigner and Bunge should not be overemphasised. The reason is that the physical content of many fundamental laws can be formulated in both ways, as laws about laws

[17] Ibid., p. 19.

[18] Bunge (1967, SRI), p. 364. See also the chapter "Laws of Laws" in Bunge (1967, SRI), p. 363 ff.

or meta-laws and as laws about systems, be they reference systems or physical systems. This can be shown by the following examples:

Relativity in Classical Mechanics

All inertial systems are equivalent. The laws of motion are invariant under change of inertial systems.

Parity in Electrodynamics

The electromagnetic interactions[19] obey parity invariance. The laws of electrodynamics (Maxwell's equations) are invariant under change of parity.

Permutational Invariance

All physical systems, which differ only by an exchange of elementary particles of the same kind (electron for electron, proton for proton, etc.) are equivalent. All laws of nature are invariant with respect to permutational change.

CPT Invariance

All phenomena of radioactive decay obey CPT invariance. All radiation laws are invariant with respect to combined symmetries of charge conjugation change of parity and time reversal.

(4) "Laws" which are not invariant (not symmetric).
To speak about laws which are not symmetric or not invariant seems to contradict what has been said in 5.2.1 and 5.3.1(1) namely that every law expresses an invariance. The answer to this problem is however rather simple: Though every law (and therefore every law of nature) expresses an invariance and may therefore be called an invariance principle (or a principle of symmetry) no law of nature is invariant or symmetric with respect to every (or arbitrary) changes or transformations.[20] Therefore it is very well possible that some law is not invariant with respect to transformation T1 but is invariant with respect to transformation T2. Thus for example Newton's laws of mechanics are not Lorentz invariant but they are Galilean invariant. Or, laws of radioactive decay are not parity invariant but they are CPT invariant. The examples could be continued along that way.

However, it will be understood from what has been said in (1) and (2) above: Something is called a law because it is invariant (symmetric) with respect to some change or transformation whereas asymmetries with respect to other changes or transformations accompany the law but are not a reason to call something a law.

[19] This does not say that in the solutions of the equations there could not be symmetry breaking (see 5(a) below).

[20] For laws of logic see Chap. 3.

(5) Invariant laws and physical systems

The relation of laws to physical systems depends on how we interpret "physical system". There are two ways which are relevant here: (a) "physical system" is understood as a material system at a certain instant of time, more precisely as a concrete state of a physical system. (b) "physical system" is understood as a material system throughout the expansion of time, more accurately as a temporally ordered set of states of that system.

According to (a) an invariant (symmetrical) law can describe, explain and predict non-symmetrical phenomena. Because a differential equation can have non-symmetrical solutions which represent the (predicted) non-symmetrical phenomena. Thus non-symmetrical phenomena (in nature) do not imply non-symmetrical laws but can be caused by non-symmetrical initial conditions. This will be discussed in more detail in the answer to the objections 2 and 3 in Sect. 5.4.2 below.

According to (b) there is a certain kind of correspondence between the laws and the systems they describe:[21] To say a law L describing the physical system S is invariant (symmetrical) under a transformation G corresponds to the fact that S does not change if the transformation G is executed on the physical system. For example: To say Newton's laws of motion (L) describing the system sun–earth (S) are Galilean invariant (G) corresponds to the fact that the system sun–earth does not change when it is transformed internally or moved inertially. Similarly with other laws describing other physical systems. Though in the present example (sun–earth) a real execution is impossible there are cases where the internal transformations and the inertial movement can be executed. Such a case is described in detail by Galileo (see Sect. 5.4.2 below).

The respective correspondence also holds between statistical laws and the system they describe: To say that the law of entropy, describing a process of approaching the macrostate of a relative equilibrium, is invariant with respect to changes of microstates corresponds to the fact that whatever the movements of the (regular or irregular) individual microprocesses are, the macrostate which is produced by averaging billions of those microprocesses has statistical regularity.

This kind of correspondence is analogous to the comparison in section (3) above between meta-laws which state properties of laws and laws about (physical) systems which state the corresponding properties of these systems. Therefore the examples of section (3) can be applied also here in an analogous way.

---

[21] From a more fundamental point of view the correspondence between laws and objects which they describe is discussed in Chap. 10.

(6) Active and passive transformations

A *laboratory system* consists of a material basis which is equipped with rods and clocks and will also be referred to as a *reference system.*[22] Assume a laboratory of this kind or a more perfectly equipped one like the ship described by Galileo (see 5.4.2 below). Then the laws, which hold in this reference system will not change, i.e. they are invariant if we change

(a) the position of the laboratory;
(b) the orientation of the laboratory (i.e. turn it on an angle);
(c) the day of observation (to the past or to the future);
(d) the state of motion of the laboratory (from rest to uniform velocity on a straight line (inertial movement) or from one inertial movement to a different one).[23]

On the other hand the laws describing the same system will also not change (stay invariant) if we change the coordinate system of the laboratory instead of changing the laboratory in the sense of (a)–(d): That means, if we change respectively

(a′) the names of the places for the location (space coordinates);
(b′) (turn) the environment around for the same angle or better: just change the names of the places in the surroundings;
(c′) (reset) the clocks in the laboratory (reference frame);
(d′) the names (numbers) of the measurement units on a straight line (x-axis) in accordance with a motion with constant velocity.

Changes of the laboratory in the sense of (a)–(d) are called *passive transformations*, whereas changes of the coordinate system of the laboratory in the sense of (a′)–(d′) are called *active transformations.*[24] The important point to notice here is this: Since the laws are invariant with respect to both, passive transformations [(a)–(d)] and active transformations [(a′)–(d′)] it follows that both transformations are equivalent. Or in other words: If the laws (of nature) describing a physical system S are invariant under transformations G then active and passive transformations in the sense of G are equivalent (with respect to S).

5.3.2 Groups of Symmetries (Invariances)

The symmetries (invariances) of contemporary physics go beyond the two traditional ones of space and time: First with respect to a rich differentiation

[22] From a more systematic point of view the concepts of *coordinate system* and *reference system* will be treated in Sect. 6.3 below.

[23] Observe that the word "change" must not be misunderstood here. All what is meant is that the laws don't care whether the laboratory is at rest or whether it moves with a constant velocity (or a different constant one) on a straight line.

[24] For more details on active and passive transformations see 6.4.7.2 (4) and Mittelstaedt (1995, KLM), p. 36 ff.

inside the two traditional ones of space and time and second with respect to new symmetries which have not been known in the tradition.[25]

(1) Permutational symmetry
Permutational symmetry means that "different individual" particles of the "same kind" are treated identical. Thus the laws and the respective physical reality described by these laws remain the same if we interchange any two electrons. The same holds for protons, neutrons, neutrinos and μ-mesons (according to the Fermi–Dirac statistics) and also for photons, π-mesons, K-mesons and gravitons (according to the Bose–Einstein statistics).

Permutation, i.e. the exchange of elementary particles of the same kind does not change the laws of physics. However, this is not necessarily also the case for a physical system or its state, respectively. There is an important difference between bosons (which are subject to the Bose–Einstein statistics) and fermions (which are subject to the Fermi–Dirac statistics and which obey Pauli's *exclusion principle*). In case of bosons the exchange of two particles does indeed not change the state of the physical (two-particle) system. This means that bosons of the same kind are treated as indistinguishable although numerically different. However, for fermions there are states, which change sign if two fermions are exchanged.

Philosophically, it may seem controversial if by some principle of individuation numerically different though indistinguishable particles are viewed as different individuals. In this connection it is interesting to look at Leibniz's principle of the identity of indiscernibles.

Leibniz's Principle of the Identity of Indiscernibles

According to Leibniz's requirement, that indiscernibility of substances implies their identity, any two individuals must differ by some *internal* properties. For

> "it is always necessary that besides the difference of time and place there should be an internal principle of distinction".[26]

Leibniz had three main reasons for his principle of identity of indiscernibles. First he thought that it follows from his principle of sufficient reason:

> "I infer from the principle of sufficient reason among other consequences, that there are not in nature two real, absolute beings indiscernible from each other; because if there were, God and nature would act without reason in ordering the one otherwise than the other".[27]

---

[25] The rough division into four kinds of symmetries is taken from Lee (1988, SAW), Appendix.

[26] Leibniz (NE) 2, 27, 1.

[27] Leibniz (GPh), 5th letter to Clark (GPh) 7, p. 393.

For Leibniz this means also that the assumption that two individuals (or substances) are completely equal cannot have a proof from (true) axioms; because the system of reasons in nature or of the right axioms is complete. Therefore his principle of sufficient reason has two parts:

> "that nothing is without reason, or that every truth has it's a priori proof".[28]
> "It is certain therefore, that all truths, even the most contingent, have an a priori proof, or some reason why they are rather than are not. And this is itself what people commonly say, that nothing happens without a cause, or that nothing is without a reason."[29]

However it is hard to see why God (or nature) would act without reason by designing two equal individuals which differ only in external (spacetime) properties. But Leibniz goes so far to claim – and this is his second reason – that from such a supposition it would follow that there would not be different individuals at all; and because of such an absurd consequence there could not be atoms such that the notion of "atom" is chimerical: "If two individuals were perfectly alike and equal and, in a word, indistinguishable in themselves, there would be no principle of individuation; and I even venture to assert that there would be no individual distinction or different individuals under this ... condition. Therefore the notion of atom is chimerical and has its origin only in the imperfect conceptions of men."[30]

The third reason and the motivation for these claims are mainly to be found in Leibniz's monadology, i.e. in his theory of substance. The simple substances or monads are spiritual units of activity, which differ from each other in two internal properties: in their activity and in the fact that each of them mirrors the universe in a different way. From this point of view the atoms of Democritus didn't seem to be genuine atoms for Leibniz.[31] Coming back to permutational symmetry the main reason for not accepting Leibniz's principle here is the criterion by laws of nature: An interchange of two elementary particles of the same kind does not affect laws, i.e. such particles are treated indistinguishable by laws (though they are numerically different). And consequently we speak of permutational symmetry.

There is, however, an important problem, which should briefly be mentioned. Is there a difference between fermions and bosons with respect to

[28] Leibniz (GPh) 2, p. 62.
[29] Leibniz (GPh) 7, p. 300 f. That Leibniz's principle of sufficient reason is a completeness principle has been elaborated in detail in Weingartner (1983, IMS) especially Chap. 2.5.2. See also Sect. 11.1 of the present book.
[30] Leibniz (NE) 2, 27, 3.
[31] This is however not the place to go into Leibniz's metaphysics any further. It suffices to understand that his "internal principle of distinction" is based on his theory of substances or monads. In addition to that also macroscopic objects (like leaves, plants or animals which he uses as examples) may have lead him to this view.

Leibniz's principle? How relevant is the fact that fermions (in contrast to bosons) obey Pauli's exclusion principle? Indeed, the exclusion principle (no two fermions of the same kind can occupy the same state) suggests a possible connection with Leibniz's principle (no two substances can be completely equal but different in number). On this ground it is sometimes claimed, in the literature, that Leibniz's principle is vindicated for fermions. For example, Pauli's exclusion principle is explicitly referred to as "Leibniz–Pauli exclusion principle" by Weyl.[32] The discussion of this point, starting with Margenau's argument that electrons violate Leibniz's principle, generally focuses on the question of what states can be attributed to the individual components of a system of "identical" fermions.[33] Finally, the discussion led to the commonly accepted result that, if quantum mechanical description is to be considered complete, there is no way of vindicating Leibniz's principle for fermions.[34] Particularly relevant, in this regard, is the impossibility of the so-called "ignorance interpretation" for a certain class of mixes states, i.e. the non-objectification theorem stating that general mixed states do not admit an ignorance interpretation.[35]

Moreover, there is another interesting and very deep consequence of Leibniz's view concerning indiscernibles.[36] It is directly connected with his objections against Newton's concept of absolute space: Leibniz pointed out against Newton that if we change the absolute positions of all material bodies while preserving their relative positions then the two states which correspond to this change are indistinguishable. Absolute position and absolute velocity do not have real significance and Newton's laws of motion cannot make a distinction between such states. This is expressed by the principle of relativity in classical mechanics which claims that all inertial systems are indistinguishable. Thus Leibniz has seen an important point here: that Newton's laws (of motion) satisfy a principle of indistinguishability of certain states and in general that they satisfy a relativity principle with respect to inertial systems. Moreover it should be noticed that taken literally Leibniz's definition of identity of indiscernibles as agreement in all properties is hardly applicable at all. The identity expressed by physical equations does not always satisfy this definition, e.g. if on the right hand sight there are observable magnitudes (for example masses, lengths, times) and on the left side not (forces) – like in Newton's second law of motion. That is "for all properties" has to be restricted in a suitable way – for example for all physical magnitudes representable by real numbers.[37] Thus a stronger principle of individuation may not be helpful

[32] Weyl (1949, PMN).

[33] Margenau (1944, Epr).

[34] More details about this debate can be found in Castellani, Mittelstaedt (2000, LPP).

[35] Busch, Mittelstaedt (1991, POQ).

[36] For a detilled discussion see Friedman (1983, FST), Chap. VI.

[37] Similarly in mathematics the right part of the equation may contain complex numbers whereas the left part does not.

here. The distinctions and differences go as far as discovered inner properties go; if new such inner properties would be discovered which hold only for a part of the particles belonging to one kind, then new differences will appear.[38]

(2) Continuous spacetime symmetries
   (a) Translation symmetry in space. This leads to three conservation principles of momentum. Unobservable: absolute place.
   (b) Rotation symmetry in space. This leads to three conservation principles of angular momentum. Unobservable: absolute direction. "Rotation" means here turning of the physical system on an angle. It does not mean that the system is rotating.
   (c) Translation symmetry of time (i.e. delay in time makes no difference). This leads to the principle of conservation of energy. Unobservable: absolute (point of) time.
   (d) Velocity symmetry: Invariance with respect to transformations between systems of inertia, thereby assuming the existence of a universal time. Unobservable: absolute velocity.

   (a)–(d) plus the explicit assumption of Euclidean space provide the full Galilean invariance (symmetry) which includes relativity with respect to inertial frames (d). This symmetry group (a)–(d) underlies Newton's theory. Observe that condition (d) was already discovered by Galileo although fully understood only by Newton. This is the important principle of relativity entering the theory of special relativity when the existence of a universal time is dropped.

   (e) Lorentz invariance: Invariance with respect to transformations between systems of inertia, without thereby assuming a universal time, i.e. one unique measure of time for all reference frames. This is the invariance of special relativity.
   (f) Invariance through the equivalence principle of general relativity: Physical laws are the same in all "free falling" (with acceleration moving) local inertial systems and agree in these systems with the laws of special relativity. This is the so-called weak equivalence principle. The underlying equivalence, which also holds in general relativity is that the inertial mass and the gravitational mass are proportional with a universal factor.

(3) Discrete symmetries
   (a) Charge conjugation symmetry or particle–antiparticle symmetry. Unobservable: absolute sign of electric charge. This symmetry is in fact not universally valid. Experiments with neutral K-mesons show that

[38] It should be mentioned that there is an extensive discussion about the problem whether the Leibniz principle is relevant at all for measurable physical quantities, since physics is concerned merely with *external* properties which are, as place and time, not sufficient for distinguishing two substances. cf. Castellani, Mittelstaedt (2000, LPP).

$K_L{}^0$ particles, though electrically neutral, decay faster into e^+ than into e^-.

(b) Parity or right–left symmetry or mirror-image symmetry. Conservation: Parity. Unobservable: absolute right or left, viz. absolute orientation in space. This symmetry is satisfied for electromagnetic effects but not completely fulfilled for radioactive decay phenomena.

(c) Time reversal symmetry or past–future symmetry. Unobservable: Absolute direction of time. No conservation principle follows. All fundamental laws of physics (of quantum mechanics and of the theory of relativity) seemed (so far) to be invariant with respect to time reversal. According to Prigogine this is a sign that the laws of physics are still incomplete since many processes are irreversible in time.[39] But time-reversal symmetry seems not to hold on the micro level.[40]

(d) CPT symmetry. This symmetry seems not to be violated by any processes known so far. This is a symmetry without conservation principle.

(4) Gauge symmetries

Gauge symmetry means that the physical world remains the same if in electrodynamics, say, the 4-vector potential A_μ is replaced by $A_\mu + \Lambda_{i\mu}$ with an arbitrary scalar function Λ or if in quantum mechanics the wave function ψ is multiplied by some phase factor. A consequence of this symmetry is the conservation of electric charge and – when applied to other phases – the conservation of hypercharge, baryon number and lepton number. Unobservable: the phase difference between two states of different charge.

(a) U1 symmetries. They lead to conservation laws of baryon number, lepton number, electric charge and hypercharge.

(b) SU2 symmetry: Isospin symmetry. This symmetry which means an interchange of proton and neutron is not completely satisfied because of the slight mass difference (0,14%) of proton and neutron.

(c) SU3 symmetry: Colour and flavour symmetry. This symmetry is the basis of quantum chromodynamics.[41]

It can be seen from the above list of symmetries or invariances that on the one hand every of the symmetries listed either leads to a particular law or is satisfied by some or by all laws. On the other hand it also becomes clear that every law is an invariance or symmetry principle: because every law is something which does not change (is stable) relative to something which

[39] cf. Prigogine (1993, TDC) and Prigogine, Stengers (1993, PZt). See Sect. 7.2.3.2 below.

[40] There is indirect evidence via CP violation and CPT symmetry. There is direct evidence shown by new experiments. cf. Schwarzschild (1999, TEO) and Hagiwara et al. (2002, PDG).

[41] That there are gauge symmetries already in classical mechanics was pointed out by Mittelstaedt (1995, KLM), p. 318 ff.

changes and since what does not change is called invariant or symmetric it follows that every law is invariant or symmetric relative to some change or transformation.[42]

An invariance (symmetry) which seems to be satisfied by all laws of nature (and by all laws of physics) is permutational symmetry. As far as we know, also the two oldest (known) symmetries – translation symmetries in space and in time – are satisfied by all laws of nature (and laws of physics). There is however the large numbers hypothesis of Dirac which says that the constant G (gravitational constant) is not really constant such that laws using G (directly or indirectly) would no more be strictly (continuous) time symmetric (invariant).[43] Since the experimental results for a change of G are so far negative it is believed that also the other continuous spacetime symmetries are satisfied by all laws of nature.

Many laws satisfy parity and charge symmetry, but not all of them do so, although CPT symmetry seems to be satisfied by all laws of nature (and of physics).[44] Time-reversal symmetry seemed to be satisfied by the fundamental laws of quantum mechanics and of the theory of relativity but is now presumably violated on the micro level. Moreover, it is not satisfied by the law of entropy and all those laws which describe phenomena where the recurrence of the state of the whole system is very improbable.[45] Of the gauge symmetries only some of U1 symmetries and the SU3 symmetries seem to be satisfied by all laws.

5.3.3 Symmetry and Symmetry-Breaking

The notion “symmetry breaking” is not only applied to laws on the one hand and to nature (or physical systems) on the other (as it was pointed in the terminological remarks at the beginning) but is used also in the sense of explicit and spontaneous symmetry breaking: If the symmetry is clearly dominant and there are some small exceptions (symmetry breaking) then one speaks of explicit symmetry breaking. Example: parity and charge violation. If there is a transition from a rather complete symmetry to a state of ordered structure one speaks of spontaneous symmetry breaking. Example: The melting charge cools down and a crystal grows. The following considerations concerning symmetry and symmetry breaking refer to both types of symmetry breaking.

For the following consideration it will be presupposed what has been said under 5.3.1(5), i.e. there will be no confusion if we speak of symmetric (invariant) laws on the one hand and of symmetric or asymmetric phenomena, physical systems, biological systems etc. on the other. In this chapter we will

[42] That this does not hold only for physical laws on which we focus here but also for laws in biology is shown convincingly by Woodward (2001, LEB).

[43] For more on that see Chaps. 8 and 9.

[44] See below to objection 5.1.4 and 7.2.3.5

[45] See below Sect. 7.2.3.2

show two things: First (a) that the demarcation between symmetry and symmetry breaking depends (is relative with respect to) several aspects. And second (b) that symmetry-breaking does not destroy symmetries but hides them.

(a) *The demarcation between symmetry and symmetry breaking depends on several aspects* The three following ones are important: (i) on the kinds of phenomena considered, (ii) on the level considered, for example macro- or micro-level or on the length of time to be observed (iii) on special quantities, like distance or temperature, etc.

  (i) If we consider electromagnetic phenomena we have parity (right–left symmetry), if we consider radiation phenomena we have symmetry breaking (of parity). Another example is Lorentz invariance with respect to electromagnetic phenomena (satisfied) and with respect to Newton's mechanics (not satisfied). Such examples could be continued.

  (ii) If we consider a litre of gas (isolated) at room temperature i.e. on the macrolevel, its macrostate is highly symmetric: translation, rotation and mirror symmetric. A snapshot of its microstate, however, will show a strong symmetry-breaking with respect to all the three symmetries of the macro state. Concepts like temperature, velocity distribution of the molecules, pressure, etc. presuppose or use the highly symmetric macro state which is the statistical average of a huge number of possibly microstates. Similar with a vacuum. If we make a "picture" of the vacuum with infinite exposure time the vacuum would appear as completely symmetrical; however if we make a "picture" with short exposure time we could see all the fluctuations and the complete symmetry would be broken. Also the microstates of a gas, observed over a long time would become a symmetric state. A further example is the cosmological principle. It says that the universe on a large scale (and in very large parts) is homogenous and isotropic, i.e. symmetric with respect to translation and rotation. But this symmetry is due to averaging over very large parts; locally the distribution of masses is very asymmetric; large empty spaces on the one hand and very high concentrations of masses on the other.

  (iii) Symmetries are dependent on the strength and the range of interactions: Unification of electromagnetic with weak interactions begins at a range of 10^{-16} cm and smaller; above that range there is no symmetry of electroweak interaction. According to the GUT scale there is no symmetry for strong and electroweak interaction above 10^{-29} cm. Another example concerning temperature is a crystal (symmetry breaking) which is produced at a certain temperature from the symmetric melting charge of iron which is rotational symmetric above the Curie temperature but becomes magnetic (symmetry-breaking) below that temperature.

(b) *Symmetry breaking does not destroy symmetries but hides them*
The examples given above suggest already the thesis that symmetry breaking on one level presupposes symmetry on a deeper and more fundamental level. This idea is expressed nicely by Lee in the following passage:

> "What is the difference between these two views: an non-symmetrical natural law? Or a non-symmetrical world? Insofar as we accept the fundamental law of nature to be immutable and permanent, while the world obviously undergoes continuous change, these two possibilities are clearly distinct from each other, though not mutually exclusive. A non-symmetrical law implies a non-symmetrical world, but not vice versa. Since we are perhaps more accustomed to a world which is somewhat skewed, it seems at least meaningful to inquire whether all the recent discoveries of symmetry violations are consistent with our fundamental physical laws being totally symmetrical."[46]

This passage points out two things: first, that although a "non-symmetrical law implies a non-symmetrical world" it does not follow that a non-symmetrical world implies a non-symmetrical law, i.e. a non-symmetrical world can be very well described by symmetrical laws plus "non-symmetrical" initial conditions.[47] Secondly, that discoveries of symmetry-breaking can be consistent with totally symmetrical fundamental physical laws which might be hidden.[48] But is there a reason to prefer the symmetry, the symmetrical law on a deeper (perhaps hidden) level over the symmetry-breaking or non-symmetrical law? We think there is, and the main reason is connected with our understanding of what a law is: A law is something general (recall condition G2 of genuine laws) and does not describe individual objects as individuals. Thus a single particular microstate of a gas, or a single crystal, or a single star – all these are symmetry-breakings – are not the concern of a law. On the other hand properties like temperature, velocity, pressure, density, etc. which enter symmetrical laws are based on averaging over a huge number of objects or single states or events. By a similar consideration concerning universality we prefer Planck's radiation law to both, that of Rayleigh and Jeans and that of Wien, since the latter hold only for the long waved or short waved part of the spectrum respectively, whereas Planck's law shows the higher symmetry. The respective more general invariance – the higher symmetry – may be hidden.

The point that symmetries (invariances) might be hidden and the asymmetry may have been caused by some initial conditions during the evolution can also be illustrated by the following examples from biology. A biological symmetry-breaking is that our heart is on the left side. But since there are few cases where children are born with the heart on the right side the biological laws do not seem to exclude such cases, i.e. the laws seem to be invariant with

[46] Lee (1988, SAW), p. 20 f.
[47] cf. The answer to objection 2 and 3 below, Sect. 5.4.2.
[48] cf. also Genz, Decker (1991, SSP), p. 357.

respect to right or left and the selection of “left” (for most cases) seems to be produced by initial conditions. Another example are the houses of snails which have the spiral turning in one direction within one species but with a few exceptions where the spiral turns the opposite way. Beans grow in a helical way of one direction. Similarly when polarised light is twisted to the right when passing through a sugar solution. A fluid may possess hidden symmetries before it crystallises. In many such cases we do not know the exact kind of initial conditions and when and how they produced the asymmetry during the evolution of the universe.

5.4 Laws, Constants, Symmetries: Answer to the Objections

5.4.1 Selection of Parameters Entering Laws (to Objection 5.1.1)

The selection of parameters and variables entering a law is not arbitrary in a twofold way: First in a direct way by successful search for the relevant parameters and variables (1). Secondly in an indirect way through limitations by constants (2) and scales (3).

(1) Successful search for the relevant parameters and variables can occur in different ways:
   (a) By considerations of structural analogy. An example for this way are the matter waves found by De Broglie by analogous transmission of the relation between wavelength and momentum to material particles.
   (b) By generalisation. When raising a law to a higher level of symmetry or invariance certain parameters disappear but deeper parameters remain. For many examples recall Sects. 5.3.2 (Groups of Symmetries) and 5.3.3 (Symmetry and Symmetry-Breaking)
   (c) By investigation on statistical correlation of variables. Quantitative measures of statistical (linear) correlation of variables provide a more precise way to judge whether two variables are mutually relevant in a given domain, i.e. whether a change in the values of one of the variables makes a difference in the values of the other.
   (d) By systematic investigation of the conjectured parameters and variables. This can be illustrated by the following example:[49]

   Consider the investigation of a spherical pendulum with small oscillations. We might start by listing what we suspect to be the relevant parameters (variables) and their dimensions: Period of oscillation T (s); length of pendulum l (cm); mass of pendulum m (g), acceleration of gravity g (cm·s^{-2}), angle of swing γ. This selection of parameters is already guided by a conjectural theoretical model which allows us to dispense (at least provisionally) with the following parameters: air (damping), air (draught), suspension

[49] This example is due to Bunge (1967, SRI), p. 322.

(assumed to be rigid), thread (assumed to be inextensible) other environment (assumed to be not disturbing). In this sense the theoretical model describes an "ideal" pendulum. The general relation among the selected parameters is $R(T,\ l,\ m,\ g,\ \gamma)$ and since we are looking for a law we are looking for an invariance under changes (of units) of one or several of these five parameters. Suppose we want to find a solution for T as a function of l, m, g and γ. Then we might proceed as follows.

First discovery: a change of the (unit of) mass does not change any of the four other parameters. Thus mass is not a relevant parameter. Result: $T = F(l,\ g,\ \gamma)$.

Second discovery: changes of l affect g if T does not change. Result: $T = f(l/g, \gamma)$. Since γ is dimensionless and the dimension l/g is T^2 the result is $T = \sqrt{l/g} \cdot f(\gamma)$.

Third discovery by experiment: $f(\gamma)$ is close to 2 π. Result: $T = 2\pi\sqrt{l/g}$. Concerning the selection of parameters and variables entering a law the above example shows this:

(i) The selection at the beginning of some systematic investigation, since it is conjectural to a high degree, may be more or less arbitrary.
(ii) In the process of further investigation the arbitrariness decreases step by step and the demarcation between relevant and irrelevant parameters (variables) becomes more and more definite.
(iii) The obtained result (relevant parameters and low-level law) shows a clear selection of the relevant parameters for the respective law. Though this result is of course relative to a certain degree of accuracy and a certain level of generality (larger oscillations or air damping make the law more complicated); but relative to that domain the selection of the parameters is not arbitrary.

Concluding section (1) it can be said that with respect to the result of the respective investigation the parameters and variables entering a law are not arbitrary.

(2) Limitations by constants[50]

Important constants enter laws of nature and thus restrict and select other parameters occurring in these laws such that they cannot be arbitrary. This can be shown by the fact that other parameters can be defined with the help of these constants and this results in a constraint on the remaining variables occurring in the law.[51] For non-relativistic quantum mechanics three constants are very important: Planck's constant h (or for many occasions more practical: $h/2\pi = \hbar = 1.055 \cdot 10^{-27}\mathrm{g} \cdot \mathrm{cm}^2 \cdot \mathrm{sec}^{-1}$), the mass of the electron $m_e = 9,11 \cdot 10^{-28}$ g and the elementary charge of the electron $e = 4.81 \cdot 10^{-10}\ \mathrm{g}^{1/2}\,\mathrm{cm}^{3/2}\,\mathrm{s}^{-1}(= 1.6 \cdot 10^{-19}$ coulomb). With

[50] For more on fundamental constants see Chap. 8.

[51] For more details concerning the subsequent considerations see Genz, Decker (1991, SSP), p. 302 ff.

the help of these three constants one can define a length (equal to the Bohr radius that is the radius of the hydrogen atom in its ground state) and a time which is related to the energy E_e of the electron in this state by $\hbar/2t = E_e$.

The above-mentioned three constants are the important constants of the non-relativistic Schrödinger equation of quantum mechanics, if hydrogen-like atoms are treated. The mass parameter is – by approximation – that of the electron mass only, because the masses of the atomic nucleus is very big in relation to that of the electron or one may take the reduced mass m of both nucleus and electron: $m = m_N \cdot m_e/(m_N + m_e)$.

Taking Maxwell's equations – which describe electric and magnetic phenomena instead of the Schrödinger equation – requires the addition of the constant velocity of light c if velocities comparable to c are permitted. As already mentioned earlier Maxwell's equations are Lorentz invariant. With the help of $\hbar, e$ and c the fine-structure constant α can be defined as $\alpha = \mathrm{e}^2/\hbar \cdot c \approx 1/137$.

For quantum mechanics combined with general relativity the three main constants are h, c (light velocity) and G (gravitational constant), where $c = 3 \cdot 10^{10}\,\mathrm{cm\ sec}^{-1}$ and $G = 6.67 \cdot 10^{-8}\,\mathrm{cm}^3\mathrm{g}^{-1}\,\mathrm{sec}^{-2}$. With the help of these three one can define the standards on the "Planck's scale", the Planck length $l_{Pl} = 1.6 \cdot 10^{-33}\,\mathrm{cm}$, the Planck mass $m_{Pl} = 2.176 \cdot 10^{-5}\,\mathrm{g}$ and the Planck time $t_{Pl} = 5.39 \cdot 10^{-44}\,\mathrm{sec}$. The standards of the Planck scale are the same for every system and are therefore invariant in a similar sense as the laws of nature.

These limitations by constants show that the selection of parameters entering a law and those with respect to which the law is invariant are not arbitrary or unlimited as claimed in objection 5.1.1; therefore its conclusion is not proved by that argument.

(3) Limitations by scales

The selection of parameters occurring in laws (of nature) are also not arbitrary with respect to scales since the laws of nature are not scale invariant. Scale invariance means that there is no natural scale; or in other words that one cannot distinguish a system from its (by some factor) enlarged copy just by its inner properties (laws). A map can be enlarged or reduced in order to be suitable. But atoms cannot be enlarged or reduced and thus laws in which basic physical constants play a role are not scale-invariant.

Scale invariance is also called affine invariance or invariance with respect to an affine group. An affine group translates points into points, straight lines into straight lines, planes into planes, cubes into cubes, balls into balls, etc. Thus it forgets distances but keeps intersection properties. Thus such a transformation – also called scale transformation – changes all lengths by the same factor a, all planes by a^2 and all volumes by a^3, if $a > 1$ or $a < 1$ respectively. A modern copying apparatus enlarges or diminishes copies, maps are diminished representations of countries or cities,

etc. But can we enlarge things of nature or art (technical constructions) for instance can animals or trees or ships or buildings be twice or three times or hundred times as large as they are? This was already a question for Galileo and he answered it in the negative:

> "From what has already been demonstrated, you can plainly see the impossibility of increasing the size of structures to vast dimensions either in art or in nature; likewise the impossibility of building ships, palaces, or temples of enormous size in such a way that their oars, yards, beams, iron bolts, and, in short, all their other parts will hold together; nor can nature produce trees of extraordinary size because the branches would break down under their own weight; so also it would be impossible to build up the bony structures of men, horses, or other animals so as to hold together and perform their normal functions if these animals were to be increased enormously in height; for this increase in height can be accomplished only by employing a material which is harder and stronger than usual, or by enlarging the size of the bones, thus changing their shape until the form and appearance of the animals suggest a monstrosity.[52]

Galileo was right to understand that we cannot enlarge or diminish reality. Even if possible – within certain limits – for things made by technical construction we cannot make an enlarged or diminished copy of an atom. The radius of the hydrogen atom (in its ground state) is fixed and about $5 \cdot 10^{-9}$ cm. Also the Avogadro number is fixed and $6.022 \cdot 10^{23}$ mol^{-1}. On the other hand the smallest distance between elementary particles in collision, which is 10^{-18} cm seems to be a technical limitation depending on the energy of our accelerators. Larger systems do not contain larger atoms but just more atoms. Since the laws of nature contain such constants they cannot be scale-invariant.[53]

Again the limitations by scales show that the selection of parameters relevant for a law are not arbitrary, since laws of nature are not scale invariant. Therefore the conclusion of objection 5.1.1 has not been proved.

5.4.2 Symmetrical Laws and Non-Symmetrical Phenomena (to Objections 5.1.2 and 5.1.3)

As it is clear from 5.3.1(5) and 5.3.3(b) asymmetries in nature (or non-symmetrical states) do not necessarily imply non-symmetrical laws, since they

[52] Galileo (DNS), p. 130 [169]. See also the examples of Feynman (1997, SNP), p. 26 f. and (1992, CPL), p. 95 f.

[53] Observe however, that some low level (non genuine) laws like the principle of Archimedes are (within modest limits) scale invariant. In his (2002, RIS), p. 120 Suppes describes the obvious non-scale-invariance of laws of nature misleadingly by the statement that the "Fundamental equations of physics are not invariant".

may be due to initial conditions. As it was pointed out there the higher symmetry (invariance) underlying an apparent asymmetry might be hidden. Thus asymmetries in nature can be described and explained by symmetrical (invariant) laws plus "non-symmetrical" initial conditions. Or: Symmetric laws describe ("produce") and predict asymmetric phenomena if the initial conditions are asymmetric. This can be substantiated by the following consideration: Galileo understood that the physical laws are the same in different inertial frames (i.e. in frames which are – relative to one another – at rest or are moving with uniform velocity). That is the physical laws are invariant (symmetric) with respect to inertial frames. The following is a nice passage from his dialogue which describes his "Gedankenexperiment":

> "*Salviatus.* Shut yourself up with some friend in the main cabin below decks on some large ship, and have with you there some flies, butterflies and other small flying animals. Have a large bowl of water with some fish in it; hang up a bottle that empties drop by drop into a wide vessel beneath it. With the ship standing still, observe carefully how the little animals fly with equal speed to all sides of the cabin. The fish swim indifferently in all directions; the drops fall into the vessel beneath; and, in throwing something to your friend you need throw it no more strongly in one direction than another, the distances being equal; jumping with your feet together, you pass equal spaces in every direction. When you have observed all these things carefully (though there is no doubt that when the ship is standing still everything must happen in this way), have the ship proceed with any speed you like, so long as the motion is uniform and not fluctuating this way and that. You will discover not the least change in all the effects named, nor could you tell from any of them whether the ship was moving or standing still."[54]

Galileo assumed that the orbits of the planets are circles. One of his reasons was probably aesthetic. At least the aesthetic reason is already present in the Pythagoreans. First they thought that the earth is spherical, not flat; secondly so are the stars. Thus Sarton says:

> "The dogma of spherical perfection and its cosmologic consequences may be considered the kernel of early Pythagorean science".[55]

Thirdly they thought that the planets are not "errant", do not wander, but they must have circular and uniform movements. For Aristotle circular movement is the most regular and perfect, and from it the best measure-unit of time can be taken.[56]

---

[54] Galileo (DWS) Second Day. cf. the discussion in Berry (1978, PCG), p. 30 ff. The reference in Berry (p. 31) to Galileo "Dialogues concerning two new sciences" is incorrect.

[55] Sarton (1966, HSc) p. 212.

[56] For more details see Sect. 6.2.2.

The other – more important reason for Galileo – might have been his understanding that the laws of motion are rotationally symmetric and therefore allow circles as its simplest solutions even if circles are not required by the laws of motion. But what if he would have thought that the orbits are exclusively determined by the rotationally symmetric laws? Then he would have been fully justified to believe in the orbits as circles. In fact *initial conditions* in addition to the laws determine the orbits. And these initial conditions may be asymmetric, may break the symmetry; i.e. determine the deviation from circles to produce ellipses, described and explained by Kepler and Newton.

That the orbit of a planet lies in one plane which contains the sun is determined by the laws, it follows from the conservation law of angular momentum. But which plane it is or better which angle this plane has with respect to, say, another star is not determined by the laws; i.e. is dependent on initial conditions. That means that the laws would allow each of the possible planes but only one as the realised one. Thus in some sense the rotationally symmetric laws produce the asymmetry (the symmetry-breaking) of just one realised plane but which one depends on the initial conditions. In this sense Barrow says:

> "Nature has not made things quite so simple. It transpires that there exist a number of almost symmetries in Nature which we had thought for a long period to be precise before very accurate experiments were carried out. More awkward still, symmetrical laws do not necessarily give rise to events which possess that same symmetry. The laws of motion do not prefer one direction in space over any other, but perch a ball symmetrically on the apex of a cone, and it will surely fall in one direction or the other. All the directions are equally probable, none has any special significance: but this symmetry will be hidden by the particular motion that results in any outcome governed by the law."[57]

The distribution of masses in the universe is also locally asymmetric, i.e. locally non-isotropic (not rotationally symmetric) and not homogenous (not translationally symmetric). This is due to asymmetric initial conditions but in accordance with the Cosmological Principle the universe as a whole (and in large parts) is homogenous and isotropic, i.e. symmetric in respect to translation and rotation. This principle is merely a hypothesis which is fairly well corroborated by the following two empirical facts:

(1) The temperature of the cosmic background radiation is almost independent of the direction of the radiation. That indicates that at least at the time when the radiation began the universe was homogenous and isotropic. And if there were later asymmetries by asymmetric conditions they did not affect the temperature of the background radiation.

[57] Barrow (1988, WwW), p. 115.

(2) The velocities of the expansion of galaxies at far distance are roughly proportional to the distance of these galaxies from the earth.

The cosmological principle is an assertion of symmetry. If it holds then the laws of nature are symmetric with respect to translation and rotation and so is the whole universe.

The conclusion of this section is: As it was already pointed out by Lee (see 5.3.3(b)) a non-symmetrical world can be described and explained by symmetrical laws plus non-symmetrical initial conditions. Therefore the conclusions in objections 5.1.2 and 5.1.3 have not been proved by these arguments.

5.4.3 Explicit Symmetry-Breaking (to Objections 5.1.4 and 5.1.5)

The violation of parity and of charge conjugation are cases of explicit symmetry breaking. "Explicit symmetry breaking" means that in the overwhelming number of cases there is symmetry (between right and left and between + and −) but in a few definite cases there is no symmetry. Concerning parity we have to notice first that it is completely satisfied by all electromagnetic phenomena. But it is violated by radioactive phenomena. That means that the laws of physics are not invariant with respect to an exchange of a physical system (of the universe) with its mirror image.[58]

From this consideration it is clear that parity conservation is not completely satisfied as it is said correctly in the objection. However, it would be incorrect to conclude that there cannot be a symmetry (invariance) on a deeper level that includes parity. And in fact the combination of charge and parity, CP conservation, is already much better satisfied than the C and P separately; and moreover CPT invariance seems to be completely satisfied.

Concerning charge conjugation, its violation was established by the decay of ${K^0}_2$ mesons and ${K^0}_L$ ions.[59] That means that the experiments (nature) are not symmetrical with respect to the plus and minus signs of electric charge.

Thus the objection is correct in the sense that charge conjugation is not always satisfied. But again from this it does not follow that there are no deeper and hidden symmetries as for example the combination of charge and parity (CP). In the CP combination of charge and parity the "big" violations which occur with respect to C and P separately are almost balanced but not completely. More accurately: all violations of C and P separately (which have not been described here) are completely neutralised in CP except one. This is the one with neural ${K^0}_L$ mesons described above. The respective experiments have been repeated many times with great care such that the result is well corroborated. Therefore there is no complete CP invariance.

[58] For a lucid discussion of parity and charge violation see Lee (1988, SAW), p. 11 ff.

[59] ${K^0}_L$ has no electric charge, no spin, no electromagnetic moment and has 1000 times the mass of electron. Though K^0_L is completely neutral electrically it decays faster into e^+ than e^-.

On the other hand in all processes known so far the invariance of the combination of charge, parity and time was always confirmed. This means that there is a close connection among the three symmetries C, P and T.[60]

The upshot is: Though there is symmetry breaking with respect to parity and charge there are higher symmetries on a deeper level like CP and above all CPT symmetry.

5.4.4 Higher and Lower Symmetries (to Objections 5.1.6 and 5.1.7)

The objections 5.1.6 and 5.1.7 point to a certain relativity of symmetry and invariance with respect to observability on one hand and with respect to the demarcation between symmetry and symmetry-breaking on the other. Concerning the first (objection 5.1.6) it has to be noticed that the term "observability" and "observable" or "non-observable" can be interpreted in two ways: First in a more subjective sense requiring observers. And in this way a law of nature cannot be dependent on assumptions about observables (in the subjective sense).

Second in a more objective sense such that "non-observable" can be replaced by "disappearing parameter" or by "cancelled parameter". Since it can be grasped from the list of symmetries given in Sect. 5.3.3 that going from one level of symmetry (invariance) to a higher level of symmetry results in a disappearance or cancelling of one parameter. Thus translation symmetry in space cancels absolute position, and moving to the combined higher symmetry of both translation plus rotation symmetry in space cancels in addition absolute direction (in space). More examples of that kind can be extracted from the list of symmetries in 5.3.2. Therefore the terms "non-observable" and "observable" in the passage of Lee (objection 5.1.6) can be replaced by "disappearing or cancelled parameter" and by "appearing or emerging parameter". The passage then reads: "Since disappearing parameters imply symmetry, any discovery of asymmetry (symmetry breaking) must imply some appearing parameter". However, interpreted in this more objective sense it does not hinder that laws of nature are dependent on appearing or disappearing parameters. Since this means nothing else than that laws of nature can have different levels of universality, i.e. different levels of symmetry or invariance.

The answer to objection 5.1.7 should be clear from what has been said in Sect. 5.3.3. First of all no law of nature is invariant with respect to every parameter or transformation. Second the demarcation between symmetry and symmetry breaking is not arbitrary. It depends on important constraints and aspects like area of application, level (macro/micro) of application, length of time, distance etc. Thirdly it has been shown there, that symmetry breaking does not destroy symmetries but rather hides them. Therefore the relativity pointed out in the objection shows that laws can be considered at different

[60] For more on CPT symmetry see Sect. 7.2.3.2(3).

levels of generality; and on each level there is some symmetry. But it does not show that the demarcation between symmetry and symmetry breaking is (arbitrarily) relative since it is bound to important constraints as mentioned above. Therefore the conclusion of the objection is not proved.

On a more general point of view one may add that the concern of this book is the investigation of laws which means nothing else than the investigation of symmetries and invariances. And as is clear from condition G2 of Chap. 2, laws are not concerned with individuals as individuals. Nevertheless, as already stressed by Plato the world (universe) is a world of changing individuals. And every individual, or individual state or singularity is a breaking of symmetries. And analogously, every solution of the differential equation is a symmetry breaking relative to that equation (law). Therefore the focus of this book is that on laws and consequently on symmetries and invariances and not on symmetry breaking and individuality.

6

Is Every Law of Nature Spacetime Invariant?

6.1 Introduction. Arguments Pro and Contra

Arguments Contra

6.1.1 If every law of nature is spacetime invariant, then for physical systems obeying these laws place is irrelevant. If every physical system is embedded in some field, then place is not irrelevant. But as it seems that every physical system is embedded in some field, at least in a gravitational field, it then follows that for physical systems obeying laws of nature, place is not irrelevant. It is, therefore, not the case that every law of nature is spacetime invariant.

6.1.2 If every law of nature is spacetime invariant, then it is also space invariant. Now, space invariance means that it is irrelevant how you orient your coordinate system. But coordinate systems are mathematical spaces and not identical with physical spaces or reference frames.

Now physical laws are only invariant with respect to physical space or to reference systems. And since some physical laws are laws of nature, it follows that there are some laws of nature, which are not space invariant in the mathematical sense. And consequently: not all laws of nature are spacetime invariant in the mathematical sense.

6.1.3 If all laws of nature are space invariant, then it is irrelevant how the physical system (obeying these laws) is oriented in space. Now, it is a theorem of mathematics that every arbitrary orientation in space can be produced by a series connection of (suitably selected) movements of translation, rotation, and mirror reflection. Thus, in order to produce some kinds of orientation, one needs mirror reflection. But not all laws of nature are invariant with respect to mirror reflection, as the violation of parity shows (cf. 5.4.3).

Therefore: not all laws of nature are space invariant, and consequently, not all laws of nature are spacetime invariant.

6.1.4 If all laws of nature are spacetime invariant, then a transition of the reference frame, or of the physical system, must be possible in the same way for both space and time. There is, however, a serious difference between space and time in this respect: A physical system (object, particle) can move both in the (spatial) direction of increasing (of values) on the x-axis and in the direction of decreasing (of values) on the x-axis; but it can move only in the direction of increasing time:

> "The difference between the two cases arises from the fact that a particle's world line can cross the $t =$ constant line only in one direction (in the direction of increasing t); it can cross the $x =$ constant line in both directions. If we replace 'line' in the last sentence by 'plane', we have the generalisation of the distinction to the actual four-dimensional universe."[1]

Therefore: not all laws of nature are spacetime invariant.

6.1.5 If all laws of nature are spacetime invariant, then they seem to be invariant with respect to time reversal. But the laws of thermodynamics (like the law of entropy) and the laws of radiation (like Planck's law of radiation) are not invariant with respect to time reversal. And, as Prigogine says: "Irreversibility and probability are objective properties."[2]

Therefore: not all laws of nature seem to be spacetime invariant.

6.1.6 If all laws of nature are spacetime invariant, then the fundamental laws of quantum mechanics (QM) and of the theory of general relativity (GR) are also spacetime invariant. If the fundamental laws of QM and GR are also invariant with respect to time reversal, then it seems that spacetime invariance also implies invariance with respect to time reversal. And consequently, if all laws of nature are spacetime invariant, they are also invariant with respect to time reversal. But invariance with respect to time reversal does not hold generally: Since CP invariance has been slightly violated by the decay of neutral K^0_{L} mesons (cf. 5.4.3 above), but CPT invariance holds universally, T has to outbalance the difference. Consequently, T-invariance cannot hold unrestrictedly.

Therefore: the laws of nature do not seem to be spacetime invariant.

6.1.7 If a law determines the geometry of spacetime, then it cannot be spacetime invariant. Now, Einstein's field equations (of general relativity) determine the geometry of spacetime as functionally dependent on the matter distributed in the universe. Moreover, Einstein's field equations are genuine laws of nature.

Therefore: not every law of nature is spacetime invariant.

[1] Wigner (1972, TEU), p. 239.

[2] Prigogine, Stengers (1993, PZt), Chap. 8 (p. 315).

Arguments Pro

6.1.8 Without understanding that something is stable while something else is changing, no law could be recognised. And thus, without any invariance principle, no law of nature could be recognised. In this sense, Wigner says:

> "The first two categories of invariance principles were always taken for granted. In fact, it may be argued that laws of nature could not have been recognised if they did not satisfy some elementary invariance principles such as those of categories *a* and *b* – if they changed from place to place, or if they were also different at different times."[3]

As a consequence, invariance – such as space and time invariance – is a necessary condition for understanding what a law of nature is. Therefore, our understanding of a law of nature is such that every law of nature presupposes an invariance principle such as space and time invariance.

6.2 Concepts of Space and Time in History

Before we proceed to our proposed answer, we shall give some important views of the history of the concepts of space and time.

6.2.1 Some Highlights of the Concept of Space and Place

1. Aristotle
   Aristotle develops mainly a concept of place[4] (topos) and makes only short remarks about space.[5] Concerning the first, he gives a definition of place, which is introduced by four desiderata (or presuppositions), and chooses one of four options for the definition. The desiderata are as follows: (1) Place is what contains that of which it is the place and is not part of that which is located in place. (2) The immediate or "proper" place of a thing is neither smaller nor greater than the thing itself. (3) Place is separable from that thing which is located. (4) Every place implies a distinction between upwards and downwards, and every body has a natural tendency to move to its own special (or "proper") place.[6]

[3] Wigner (1967, SRf), p. 43. The two catergories of invariance principles *a* and *b* which Wigner mentions some lines before are translation in Euclidean space and translation in time.

[4] This is to be found mainly in his (Phys), book IV, with additions in (Heav), book II.

[5] (Cat) 5a.

[6] Aristotle (Phys), 210b35. cf. Jammer (1954, CSP), p. 16, and Thomas Aquinas (CAP), IV, Lecture 5 (445).

As a definition, he proposes then that (the immediate or proper) place of the body is the immobile inner surface of that which contains it (i.e. of the container).[7] That place is immobile is connected with (1) and (3), from which it also follows that a body is replaceable by another one at the same place. (2) indicates that Aristotle accepts that the immediate or proper place can be put into a larger context. Thus, he says: You are on the place, on the earth, in the air, in the universe.

> "To begin with, then, the phenomenon of *'replacement'* seems at once to prove the independent existence of the 'place' from which – as if from a vessel – water, for instance, has gone out, and into which air has come, and which some other body yet may occupy in its turn; for the place itself is thus revealed as something different from each and all of its changing contents."[8]

Condition (4) shows that Aristotle had already anticipated a basic idea of Mach and Einstein: Space, in the mathematical sense, has no distinguished places according to Aristotle, such places being relational to each other. This also happens, for him, in the void or absolute vacuum. But falling bodies have a definite direction, a preference distinguished by nature:

> "... whereas in Nature each of these directions is distinct and stable independently of us. 'Up' or 'above' always indicates the 'whither' to which things buoyant tend; and so too 'down' or 'below' always indicates the 'whither' to which weighty and earthy matters tend, and does not change with circumstance; and this shows that 'above' and 'below' not only indicate definite and distinct localities, directions and positions, but also produce distinct effects."[9]

As Barbour puts it: "The parallel with Mach and indeed specifically Mach's Principle becomes almost complete when Aristotle argues from the undoubted existence of the falling of bodies that they must be falling to a definite place."[10]

Concerning space, Aristotle understood it as finite, since it is an accidental property of matter, which is itself finite. As he says in the Categories, space is a continuous quantity, and Duhem seems to be right to interpret it as the sum total of all places occupied by bodies.[11] But what is important with respect to modern cosmology is that there is not anything "outside" the finite space or "outside" the finite universe, according to Aristotle. The term "outside" has no meaning here. The comparison, therefore, of this

[7] Aristotle (Phys), 212a20.
[8] Ibid. 208a; 209af.
[9] Aristotle (Phys), IV, 208b.
[10] Barbour (1989, ARM), p. 78. cf. also Jammer (1954, CSP), p. 17ff, who points to the interpretation by a dynamical field structure.
[11] Duhem (1913, SdM), Vol. I, p. 197.

conception with Einstein's "spherical space" in early relativistic cosmology seems to be well justified.[12] And as Barbour correctly points out:

> "We cannot conclude this section without commenting on the remarkable similarity of the closed universes constructed by Aristotle and, more than two thousand years later, by Einstein. Both were spatially spherical and infinite in both temporal directions."[13]

2. Thomas Aquinas
   There is a clear passage in Thomas Aquinas, which shows that, according to him, space and place are bound to the world (of material bodies). The passage occurs in the answer to an objection, which claims the necessity of a vacuum container for the world before the world existed:

   > "Objection 4: Further, a vacuum is where there is not a body, but there might be. But if the world began to exist, there was first no body where the body of the world now is; and yet it could be there, otherwise it would not be there now. Therefore before the world there was a vacuum; which is impossible."
   >
   > "Reply to Objection 4: The notion of a vacuum is not only *in which is nothing*, but also implies a space capable of holding a body and in which there is not a body as appears from Aristotle (Phys) IV 208b26. Whereas we hold that there was no place or space before the world was."[14]

3. Newton
   The most famous places where Newton speaks about absolute space are in the Scholium of the *Principia* (book I) and in *De Gravitatione*. *De Gravitatione* was written earlier,[15] and we see here an opinion, which is similar to Aristotle and to Thomas Aquinas in that space is postulated when being is postulated:[16]

   > "Space is a disposition of being *qua* being. No being exists or can exist which is not related to space in some way. God is everywhere, created minds are somewhere, and body is in the space that it occupies; and whatever is neither everywhere nor anywhere does

---

[12] cf. Jammer (1954, CSP), p. 20ff.
[13] Barbour (1989, ARM), p. 92.
[14] Thomas Aquinas (STh), I, 46, 1 ad 4. That also time is bound to the world, according to Aquinas, and that the word "before" (the world) makes no more sense than the word "outside" (the world) will be seen below.
[15] The exact date is controversial. Hall and Hall date it 1664–66, Whiteside 1670–73, some others close to the *Principia*.
[16] Though neither Aristotle nor Thomas Aquinas would claim this for every being or as a disposition of being qua being; they would exclude God, who according to Aquinas creates finite space by creating a finite material world. For both, God is immaterial, neither in space nor at some place, but can be "everywhere" in his creation (world) through his knowledge and will. In this connection see also Newton (Opt) query 28, which says that *infinite space* is the sensorium of God.

> not exist. And hence it follows that space is an effect arising from the first existence of being because when any being is postulated, space is postulated."[17]

With the concepts of absolute space (and time), Newton departs not only from the philosophers Aristotle, Thomas Aquinas and Descartes, but also from Copernicus and Kepler. From all the five, Newton departs in the sense that they have a concept of motion which is based on matter – something empirical – whereas Newton's concept is based on absolute space and absolute time, i.e. non-empirical concepts. Descartes, especially, relates motion to other bodies in the vicinity: Motion, for Descartes, is a translation from the vicinity of other contiguous parts. For Newton, this kind of "relativism" is untenable and absurd.

> "For unless it is conceded that there can be a single physical motion of any body, and that the rest of its changes of relation and position with respect to other bodies are so many external designations, it follows that the Earth (for example) endeavours to recede from the centre of the Sun on account of a motion relative to the fixed stars, and endeavours the less to recede on account of a lesser motion relative to Saturn and the aetherical orbit in which it is carried, and still less relative to Jupiter and the swirling aether which occasions its orbit, and also less relative to Mars and its aetherical orbit. Since all these endeavours and non-endeavours cannot absolutely agree, it is rather to be said that only the motion which causes the Earth to endeavour to recede from the Sun is to be declared the Earth's natural and absolute motion. Its translations relative to external bodies are but external designations."[18]

As Barbour describes this view, Newton's passionate belief was that (for every singular body) there must be *one motion* that is *true, absolute,* and *proper.*[19] This absolute motion Newton related then in his *Principia* to the unmoveable absolute space:

> "Absolute space, in its own nature, without relation to anything external, remains always similar and immovable. Relative space is some moveable dimension or measure of the absolute spaces; which our senses determine by its position to bodies, and which is commonly taken for immovable space; such is the dimension of a subterraneous, aerial or celestial space, determined by its position in respect of the earth."[20]

---

[17] Newton (Grav), p. 136.
[18] Ibid. p. 127.
[19] Barbour (1989, ARM), p. 614.
[20] Newtond (Princ), I, Scholium. A forerunner of Newton in the early Middle Ages was Philoponus, who attacked the concept of space in Aristotle, claiming that space is a given interval, measurable in three dimensions, and incorporal. See Barbour (1989, ARM), p. 91.

The critics to Newton's search and claim that there must be one true motion (for every body) pointed out that the difficulty, here, is to show how these absolute motions could be deduced from, or at least confirmed by, observed relative motions. In this respect, it is worth mentioning that Lagrange provided a solution for that with respect to the three body problem.[21] Lagrange showed that Newton's equations can be rewritten in such a form that they contained, besides the time coordinate, only relative space coordinates (for the triangle formed by three bodies which interact gravitationally). In addition, Lagrange showed how to calculate, with the help of the relative quantities, the coordinates in absolute space, i.e. in the centre of mass of the reference frame.

4. Laplace
   We have already seen that the concept of space in Aristotle had certain similarities with that of Mach and Einstein. But, also Laplace shows a striking anticipation of Einstein's curved space and even of one of the most important cosmological predications of general relativity: that of black holes. And more than that, Laplace's reason for it is an anticipation of the same reason, which was also important for Einstein: The influence of gravitation on the propagation of light. In his "Exposition du système du monde" of 1795, he makes quantitative statements about the influence of gravitation on light, on the basis of Newton's theory of gravitation. He also says there that, because of this influence, a celestial body of sufficient size could not send light rays and would be non-observable for us. Later (1799), he tried to give a mathematical proof for this claim (i.e. for the existence of black holes).[22]
5. Kant
   It is usually thought that Kant interpreted Newton's concept of absolute space – with the help of his "Copernican Revolution" – as "Anschauungsform a priori", i.e. as an a priori intuition (condition), under which sense experience can possibly operate. This is correct of Kant's conception of space in his *Critique of Pure Reason* (and in the Prolegomena). It should, however, be mentioned that Kant, in his pre-critical writings, had a completely different concept of space, which was relative in character:

   > "Now I begin to see that I lack something in the expression of motion and rest...I should never say, a body is at rest, without adding with regard to what it is at rest, and never say that it moves without at the same time naming the objects with regard to which

[21] Lagrange (1772, EPT). A detailed account is given by Dziobek (1888, MTP). cf. Barbour (2001, GCB), p. 200.

[22] This proof appeared in the *Abhandlungen der Allgemeinen Geographischen Ephemeriden-Gelehrtengesellschaft*, ed. by F.X. von Zach, Weimar. cf. Schmutzer (1996, RTA), p. 104. Engl. Trans. in: Hawking, Ellis (1973, LSS), Appendix. Schmutzer gives also a quotation of Soldner, who (in 1801) describes the deviation of light rays by the attraction of celestial bodies as a hyperbolic trajectory, with the concave side on the attracting body.

> it changes its relation. If I wish to imagine also a mathematical space free from all creatures as a receptacle of bodies, this would still not help me. For by what should I distinguish the parts of the same and the different places, which are occupied by nothing corporal?"[23]

> "...it is therefore impossible that a body would move towards another one which is at absolute rest."[24]

Ten years later, Kant had already changed his view and tried to find an evident proof for the existence and reality of absolute space independent of the existence of matter:

> "My aim in this treatise is to investigate whether there is not to be found in the judgements of extension, such as are contained in geometry, an evident proof that space has a reality of its own, independent of the existence of all matter, and indeed as the first ground of the possibility of the compositeness of matter."[25]

In his *Critique of Pure Reason*, Kant characterises space by four conditions:[26] (1) Space is not an empirical notion which has been derived from external experience; (2) Space is a necessary perception a priori or a condition for the possibility of all external perceptions or a subjective condition of sensation, because we cannot imagine that there is not space, but only that there are no objects found in space; (3) Space is no discursive or general notion but pure intuition; (4) Space is conceived as an infinite magnitude. Compared with the simple description in (NLB), the above characterisation of space contains many obscure components, which come from sacrificing the earlier view for a view bound to transcendental idealism and its a priori conditions for the possibility of experience at all.

6. Maxwell and Mach

Before we give some essential points of Mach, we also want to show that Maxwell abandoned absolute space (and absolute time):

> "Absolute space is conceived as remaining always similar to itself and immovable... But as there is nothing to distinguish one portion of time from another except the different events, which occur in them, so there is nothing to distinguish one part of space from

[23] Kant (1758, NLB), p. 3. In: (1912, KGS), Vol. 2, p. 13–25. The translation is due to Jammer (1954, CSP), p. 130. Jammer is one among the few to mention this development in Kant.

[24] Ibid. p. 5.

[25] Kant (1768, EGU). The translation is due to Jammer (1954, CSP), p. 130. The change of Kant's view seems to begin with this essay. The essay of 1763 "Versuch, den Begriff der negativen Grösse in die Weltweisheit einzuführen" does not contain a claim about absolute space (although a reference to Euler).

[26] The points (1)–(3) are contained in the first edition 1781, point (4) only in the second (1787): A24f, B38f and A26, B42.

> another except its relation to the place of material bodies. We cannot describe the time of an event except by reference to some other event, or the place of a body except by reference to some other body. All our knowledge, both of time and place, is essentially relative."[27]

But, Mach's criticism of Newton's concept of absolute space (and of absolute motion) in his Mechanics can be viewed as the earliest formulation of the principle of general relativity.[28] An important passage is this:

> "Für mich gibt es überhaupt nur eine relative Bewegung und ich kann darin einen Unterschied zwischen Rotation und Translation nicht machen. Dreht sich ein Körper relativ gegen den Fixsternhimmel, so treten Fliehkräfte auf, dreht er sich relativ gegen einen anderen Körper, nicht aber gegen den Fixsternhimmel, so fehlen die Fliehkräfte. Ich habe nichts dagegen, wenn man die erstere Rotation eine absolute nennt, wenn man nur nicht vergisst, dass dies nichts anderes heißt, als eine relative Drehung gegen den Fixsternhimmel. Können wir vielleicht das Wasserglas Newtons festhalten, den Fixsternhimmel dagegen rotieren, und das Fehlen der Fliehkräfte nun nachweisen? Der Versuch ist nicht ausführbar, der Gedanke überhaupt sinnlos, da beide Fälle sinnlich voneinander nicht zu unterscheiden sind. Ich halte demnach beide Fälle für denselben Fall und die Newtonsche Unterscheidung für eine Illusion."[29]

The observation that the "absolute motion" of the fluid in the bucket experiment is also merely a relative motion of the water in the pail with respect to the far distant masses of the universe, is the origin of "Mach's principle". This principle, which was formulated by Einstein and not by Mach, states that the local inertial forces are caused by the relative motion of the masses in the universe, and that this effect is induced by the gravitational field of the large scale cosmic mass distribution. Hence, Mach's principle should be provable within the framework of a theory of gravitation. However, in Einstein's general relativity, Mach's principle is not satisfied. The theory provides many vacuum solutions (without masses) with inertial forces.

---

[27] Maxwell (1991, MaM), Sect. 18.

[28] See Jammer (1954, CSP), p. 141, and Wien (1921, RTh), p. 31.

[29] This passage does not occur in McCormack's translation, because it is from the fourth edition (1901) and was omitted later (which was recognised by Jammer). In this respect, it should be mentioned that the two declarations against the theory of relativity in the prefaces of Mach's *Optik* (finished 1913, edited 1921 by his son Ludwig) and the 9th edition of his *Mechanik* (1933) are very probably forgery (by his son Ludwig), as Gereon Wolters has shown in a painstaking analysis (1987, MER).

6.2.2 Some Highlights of the Concept of Time

1. Aristotle

   The conception of time in Aristotle[30] can be characterised by three main features: (1) By its relation to change; (2) by the definition of time given by Aristotle, and (3) by a heavenly standard as a unit for time measurement.

   (1) The relation of time to change is twofold: (a) on the one hand, time is bound to change such that if there is no change or movement there is no time:

   > "Time cannot be disconnected from change; for when we experience no changes of consciousness ... no time seems to have passed ... Since, then, we are not aware of time when we do not distinguish any change ... it is clear that time cannot be disconnected from motion and change."[31]

   (b) On the other hand, time is independent of change and movement, in the sense that there can be faster and slower movement but not faster or slower time: "And further, all changes may be faster or slower, but not so time; for fast and slow are defined by time, "faster" being more change in less time, and "slower" less in more."[32]

   (2) Aristotle gives the definition of time as follows: "When we perceive a distinct before and after, then we speak of time; for this is just what time is, *the calculable measure or dimension of motion with respect to before-and-afterness*."[33] Before he gives this definition, and subsequent to it, he discusses two main features of it: (a) Since motion is from some place to another, and since measures and magnitudes, with respect to these places, are continuous, motion must be continuous. And since time is a measure of motion, with respect to before and afterwards, time is also continuous. (b) We apprehend time if we count different "nows" in a process of succession such that one is before and the other is after: "Time, then, is the dimension of movement in its before-and-afterness, and is continuous."[34] Thomas Aquinas, in his commentary on Aristotle's *Physics*, explains this as follows:

   > "Therefore, when we sense one 'now' and do not discern in motion a before and after, or when we discern in motion a before and after but we take the same 'now' as the end of the before and the beginning of the after, then it does not seem that time passes, for there is no motion. But when we take a before and after and number them, then we say that time

[30] cf. Barbour (1989, ARM), p. 93ff, for a lucid exposition.
[31] Aristotle (Phys), IV, 218b21.
[32] Ibid. 218b15.
[33] Ibid. 219b1.
[34] Ibid. 220a25.

passes. *This is so because time is nothing else than the number of motion in respect to before and after.* For we perceive time, as was said, when we number the before and after in motion."[35]

(3) In search for a standard unit for time measurement, Aristotle proposes a heavenly unit and tries to justify it by several points, three of which seem to be of special importance.

(a) First, a unit for time can only be taken from such a motion or change which is most basic, regular and uniform. Of the three types of motion or change, i.e. change of increase and decrease, alteration and change of place (or local motion), he tries to show that local motion is first and most basic, since the two other types presuppose local motion.[36]

(b) Further, he tries to show that circular motion is the most regular and uniform of local motion; and since local motion is continuous, circular motion is also continuous. Now, among circular motions the daily motion – in his understanding, the revolving of the firmament with the period of the day, in today's understanding, the rotation of the earth – is the most suitable for a unit for time measurement of the other motions.[37] Although revolution is not identical with time, since a part of the revolution is not a revolution, whereas parts of time are time,[38] it can serve as a unit for time measurement.

> "And now, keeping locomotion and especially rotation in mind, note that everything is counted by some unit of like nature to itself – monads monad by monad, for instance, and horses horse by horse – and so likewise time by some finite unit of time. But as we have said, motion and time mutually determine each other quantitatively; and that because the standard of time established by the motion we select is the quantitative measure both of that motion and of time. If, then, the standard once fixed measures all dimensionality of its own order, a uniform rotation will be the best standard, since it is easiest to count."[39]

(c) In his Book VIII of *Physics*, Aristotle continues to describe important features of his concept of time. Among these is his claim

[35] Thomas Aquinas (CAP), IV, 580.

[36] Aristotle (Phys), VIII, 260a27, Barbour and Barnes translate the change of place as "locomotion". Barbour has a problem of interpretation (p. 98), though Aristotle's distinction of the three types seems to be quite simple. Barbour does not seem to have incorporated Book VIII which contains an important complementation.

[37] Ibid. IV, 223b18.

[38] Ibid. IV, 218b2.

[39] Ibid. 223b13.

that the first local motion (locomotion) is eternally (infinitely) continuous. In this respect, one has to know that, according to Aristotle, the world (universe) is infinite in time (duration), i.e. has (at least with respect to prime matter) no beginning or end (cf. Sect. 6.2.1.1 about Aristotle above). Now, the first circular motion, which is the most perfect motion, is moved by the first mover (by the Aristotelean God). No other local motion other than circular motion is infinitely continuous.[40] This property of the first circular motion, that it is infinitely (eternally) continuous, i.e. going on forever in a uniform and regular way, is indeed interesting, since it suggests interpreting it in modern physical terms: as the movement of inertial systems.

2. Thomas Aquinas
   Thomas Aquinas' theory of time can be characterised by three points: First, by his definition of time, second, by its consequences, and third, by the comparison of time with eternity.
   (1) The definition of time Thomas Aquinas gives in several places in his writings is mainly that of Aristotle: Time is the measure – expressed in numbers – of change or movement with respect to before and after:

   > "Just as we derive our knowledge of simple things from composite ones so we derive our knowledge of eternity from time, which is *the measure of before and after in change.* For in all change there is successiveness, one part coming after another, and from our numbering antecedent and consequent parts of change there arises the notion of time, which is simply the numberedness of before and after in change."[41]

   (2) From this definition we can deduce several features of time, according to Thomas Aquinas:
      (a) Time is a measure, expressed in numbers, for the successiveness of change. Now, since this successiveness is continuous, time is a continuous quantity.
      (b) Time is bound to change and consequently to the changing world. Without change, there is no before and after, and consequently, no time. Before and after, or past and future, are necessary conditions for time: "Now something that lacks change and never varies its mode of existence will not display a before and after."[42] Moreover, there was no time "before" the world, but God created time by creating a changing world:

      > "Those who would say that the world was eternal, would say that the world was made by God from nothing, not that it was made after nothing, according to what we

[40] Ibid. VIII, 260a20f and 264b29.
[41] Thomas Aquinas (STh), I, 10,1.
[42] Ibid.

> understand by the word creation, but that it was not made from anything."[43]
>
> "Things are said to be created in the beginning of time, not as if the beginning of time were a measure of creation, but because together with time heaven and earth were created."[44]

(c) Since this world is finite in time, according to Thomas Aquinas, and since time is a measure of change of this world, time is finite, too. Although this cannot, however, be proved with a demonstration, since a demonstration is a proof which uses universal laws (of nature) about this world. But laws abstract from hic and nunc (place and point of time) such that no point of time, neither of beginning nor of end of the world, could be deduced from (or with the help of) universal laws:

> "The reason is this: the world considered in itself offers no grounds for demonstrating that it was once all new. For the principle of demonstration is that what something is. Now the general concept of the specific nature abstracts from here (hic) and now (nunc); whence it is said that universals are everywhere and always."[45]

(3) Time is different from eternity with respect to two features: (a) Time is limited and finite like the world is limited and finite, having a beginning and an end, since time is bound to this world. Eternity, on the other hand, is not limited but interminable; it has neither a beginning nor an end. (b) Time has an order of succession, a before and after in a process of change. Eternity has no succession, being simultaneously whole.

> "So just as numbering before and after in change produces the notion of time, so awareness of invariability in something altogether free from change produces the notion of eternity. A further point: time is said by Aristotle to measure things that begin and end in time, and that is because you can always find a beginning and an end in changing things. But things altogether unchangeable can no more have a beginning than show successiveness.
>
> Two things then characterise eternity: firstly, things existing in eternity are *endless*, lacking both beginning and end (for both may be called *ends*); and secondly, eternity itself exists as a *simultaneous whole*, lacking successiveness."[46]

[43] Thomas Aquinas (STh), I, 46, 2 ad 2.
[44] Ibid. 46, 3 ad 1.
[45] Ibid. 46, 2. Recall the discussion of this passage in the foregoing Chap. 5. note 11.
[46] (STh), I, 10,1.

3. Newton

   The place where Newton speaks of absolute time is the Scholium of the *Principia*. Here are two important passages:

   > "Absolute, true, and mathematical time, of itself, and from its own nature, flows equably without relation to anything external, and by another name is called duration: relative, apparent, and common time, is some sensible and external (whether accurate or unequable) measure of duration by the means of motion, which is commonly used instead of true time; such as an hour, a day, a month, a year."[47]

   > "Absolute time, in astronomy, is distinguished from relative, by the equation or correction of the apparent time. For the natural days are truly unequal, though they are commonly considered as equal, and used for a measure of time; astronomers correct this inequality that they may measure the celestial motions by a more accurate time. It may be that there is no such thing as an equable motion, whereby time may be accurately measured. All motions may be accelerated and retarded, but the flowing of absolute time is not liable to any change."[48]

   Concerning time, Newton wanted to avoid "that it may be, that there is no such thing as an equable motion, whereby time may be accurately measured" (see citation above); i.e. that there must be a more basic measurement of time than the one provided by the rotation of the earth relative to the stars or a similar periodic motion of another planet. In contradistinction to absolute space, absolute time, for Newton, is still in some way connected with experience: it is abstracted from by the astronomers as a "correction of the apparent time" (see citation above from the Scholium). He was searching for a genuine referential basis of observable motion and for an explanation of observable motion with the help of unobservable and absolute space and time. In this respect, he follows in the tradition of the great aim of Greek science and philosophy: To describe and explain the visible (observable), concrete, particular, changing, material world by non-visible (non-observable) abstract, universal, non-changing and immaterial principles. It does not seem that Newton ever mentions the possibility of taking the totality of bodies in the material universe – in contradistinction to a particular other body – as a frame of reference for defining motion of one body in this universe, whereas Copernicus and Kepler defined motion relative to the fixed stars (which for them were really fixed). Barbour[49] gives three reasons for Newton's adherence to the absolute concepts: His a priori rationalistic concept of space "which he was very loath to abandon"; his concentration on the demolition of Descartes' definition of motion; and

[47] Newtond (Princ), Scholium.
[48] Ibid. Scholium IV.
[49] Barbour (1989, ARM), p. 636.

his insight that all the bodies in the universe are in motion relative to each other.

Kant did not have anything new to say about time. In the four conditions for space in his Critique, he repeats his concept of time as an a priori intuition.[50] He adds the further condition that, because of the necessity provided by the a priori intuition, apodictic principles or axioms of time can be formulated. He gives as examples, that time has one dimension, and that different points of time are in succession after one another. But such features – of course without claiming a priori intuitions – are already dealt with, in much more detail, in Aristotle's *Physics*.

Maxwell's view on the relativity of time is well expressed in the passage from *Matter and Motion* cited above (6.3.1(5)). This point is stressed again when Maxwell describes a method of defining equal intervals of time.[51]

4. Mach

Mach asked the critical question whether Newton's idea concerning absolute time and its consequences could somehow be empirically tested. He came to the conclusion that absolute time cannot be used as a reference frame, at least not with any empirical meaning. On the other hand, the change or movement in time of some body can only be measured with respect to another body. More accurately, this means that by measuring time we have to compare the change of some body with a concrete clock, which is nothing but a physical process. Of course, no such process would be suitable. Remember that Aristotle had already proposed the daily rotation of the earth (in his understanding: of the stars) as a concrete clock, since circular motion was most regular to him. Time was also, for Aristotle, the measure of some concrete movement and change; and this is what Mach pointed out against Newton's idealisation of absolute time:

> "This absolute time cannot be measured with the help of any movement, it has therefore no practical and also no scientific value; nobody is justified to claim that he has knowledge of it and it is therefore a useless "metaphysical" concept."[52]

Mach also considers the rotation of the earth as a more objective measure for a time scale, since our physiological time is dependent on subjective changes, and nature does not signify an unambiguous measure. An interesting comparison between the search for an objective time scale and that for an objective scale for temperature leads Mach to the conclusion that the mutual similarities show no ultimate objectivity in both cases: All that we have is a comparison or abstraction from different scales either to an average scale or to a selection of a most suitable scale among similar ones.

[50] cf. Kant (1787, CPR), B46f.

[51] Ibid. Sect. 43.

[52] Mach (1933, MEC), p. 217.

However, there is an important dissimilarity here: It is possible to define an objective scale of temperature, based on the absolute temperature (zero kelvin), which is independent from material (used in the instruments), with the help of the efficiency of a Carnot process. This is not possible in the same way for a time scale.

Mach still did not give up his search for a concept of absolute time which was such that it could be given an empirical meaning by connecting it with a global process of the universe. He was searching for something analogous to Mach's principle with respect to space. And his proposal for such a measure of absolute time is the entropy of the whole universe.[53] Although he points out critically that the empirical meaning gained that way is hardly comprehensible.

5. Poincaré

   Similar to Ernst Mach, Henry Poincaré also realised that Newton's absolute and universal time does not exist as an experimental quantity which is measurable with well defined clocks. In this situation, the question arises as to whether we can define a metric of time by real physical processes, and whether this time is a convenient tool for formulating laws of physics. In particular, Poincaré formulates two open questions:[54]

   (i) How can we find a realisable and convenient measure of time which allows for the comparison of the duration of two subsequent processes at the same place?
   (ii) Given a realisable measure of time in the sense of (i), how can we compare the time values of two events which occur at far distant places?

The first problem had already been recognised by Newton, who pointed out that astronomers measure time intervals with respect to the motion of the sun, the planets, and the stars, which are, however, not exact clocks for the absolute time: "For the natural days are truly unequal, though they are commonly considered as equal and used for a measure of time; astronomers correct this inequality that they may measure the celestial motions by a more accurate time."[55] This is, however, a very difficult task, since "It may be that there is no such thing as an equable motion, whereby time may be accurately measured." On the basis of these considerations, Poincaré tried to find a most general criterion according to which we can uniquely define time as a measurable quantity. Obviously, we can use a given measure of time for describing the celestial motion, which will then turn out to be either uniform or accelerated. Or we could use a particular celestial motion, which we assume to be uniform, and define the measure of time by this motion. Both ways will not lead to a metric of time which is an intrinsic measure of nature and at the same time

[53] Mach, ibid. p. 319.

[54] H. Poincaré, *La mesure du temps*, Revue de métaphysique et de moral, t. VI, pp. 1–13 (1898), Chap. V.

[55] I. Newton, *Principia*, ed. F. Cajori, Berkeley, University of California Press, 1947, pp. 7–8.

experimentally well defined. Hence, Poincaré formulated a principle according to which Newton's "astronomers" could have corrected and improved the measure of time, and which can also be directly applied to physical processes: "Time must be defined in such a way that the equations of motion in classical mechanics are as simple as possible."

This principle, which is not trivial, can be illustrated in the following way.[56] If t is Newton's absolute time, then the equation of motion (in one spatial coordinate) assumes the well-known form

$$m\frac{d^2x}{dt^2} = K \tag{1}$$

If we introduce a new measure of time by $\vartheta = f(t)$, where $f(t)$ is a bijective function, then the equation of motion reads

$$m\frac{d^2x}{d\vartheta^2} = \frac{1}{(f'(t)^2)}K - m\frac{f''(t)}{(f'(t))^2}\frac{dx}{d\vartheta} \tag{2}$$

On the right hand side of this equation we find two terms, the external force K multiplied with a time dependent factor, and a new term that corresponds to a velocity dependent and a time dependent inertial force. Also without explicitly formulating a principle of simplicity, it is obvious that (1) is simpler than (2). Hence, the measure of time used in physics and astronomy seems to be chosen according to the *conventional* postulate formulated by Poincaré, that Newton's equation of motion assumes its simplest form (1).

Even if a metric of time is established in the described way, the problem arises how simultaneity of two clocks $C(x)$ and $C(x')$ at different places $x \neq x'$ can be defined. This is Poincaré's second question, and he illustrates it by the following example:[57]

In 1572 Tycho Brahe observed a nova explosion at the coordinates $\alpha = 0^h 20^m, 6; \delta = +63°44'$ in a distance of $D = 1.910^{15}$ km. If we assume isotropy of the velocity of light, i.e. equality of the velocities c^+ and c^- in both directions, then the event of the nova explosion occurred almost 200 years before its observation.

If t_O is the time of the observation which is made at the place x_O, and t_N is the time of the nova explosion at the place x_N, then $D = x_N - x_O = 1.910^{15}$ km is the distance between the nova and the observer, which is directly measurable. Hence, we have the simple relation

$$t_N = \frac{D}{c^+} \tag{3}$$

where

$$c^+ = \frac{x_N - x_O}{t_N - t_O} \tag{4}$$

[56] Mittelstaedt (1989, ZBP), pp. 28–32.

[57] Poincaré, ibid. Chap. VI.

is the velocity of a light signal which proceeds from the nova to the observer. However, there is no independent method for measuring the one-way velocity of light c^+, since for its measurement we would need two already synchronised clocks $U(x_N)$ and $U(x_O)$ at places x_N and x_O, respectively. Hence, the following problem arises: If we assume some value t_N for the nova explosion, then we can determine the value c^+ of the velocity of light according to (4). If we assume a certain value c^+ for the velocity of light, then we can determine the time value t_N of the nova explosion according to (3). In any case, we have to make a stipulation by convention either about the distant simultaneity of the clocks $C(x_O)$ and $C(x_N)$ or about the velocities c^+ and c^- in both directions. The simplest convention for distant simultaneity is the assumption that the velocity of light is isotropic, i.e.

$$c^+ = -c^- = c \tag{5}$$

where c is the two-ways velocity of light in vacuum, which can be measured with one single clock. This convention (5) for distant simultaneity was later called *Einstein synchronisation.*[58]

6.3 The Concept of Spacetime

We understand by "spacetime" – as usual – a four dimensional differentiable manifold, i.e. a set of points tied together continuously and differentiable. This means that spacetime has two characteristics:

(1) It has a topology with a definition of neighbourhood.
(2) It is coordinatisable by the set of quadruples of real numbers R^4 if there is a homomorphism $K : M_4\ R^4$ which is called a coordinate system.

6.3.1 Coordinate System and Reference System

A *coordinate system* K is given by a function $K : M_4 R^4$, which maps the four-dimensional Minkowskian Spacetime M_4 to the quadruples R^4 of real numbers. If in a *coordinate system* the space coordinates x of the material elements of the reference system are constant in time, then the *coordinate system* is said to be *comoving*. Within a given reference system there are infinitely many comoving coordinate systems.[59]

[58] cf. Mittelstaedt, ibid. p. 44.

[59] For a most general treatment of coordinate systems see Hawking, Ellis (1973, LSS). Note that the expression "comoving" is not predicated of the pure mathematical coordinate system as such. Pure mathematical structures, since they are conceptual entities, cannot move. This expression is applied to coordinate systems only insofar as they are conjoint to reference systems which have a physical (material) basis. Therefore, reference systems can move.

A *reference system* or a *laboratory system* consists of a material basis which is equipped with rods and clocks. The simplest case is a rigid reference system. Here, all parts of the system are at rest relatively to each other. In the general case, the material parts of the reference system have velocities relative to other parts of the system. Hence, in this case, a reference system is characterised by a velocity field $v(x,t)$, which attributes a velocity v to a point (x,t) in spacetime.

Although coordinate systems (as mathematical structures) – as long as they are consistent – can be constructed in different ways, as soon as they are applicable to physical systems, they have to obey certain constraints. Such constraints are necessary to preserve the causal structure of Minkowskian Spacetime. As Schrödinger[60] pointed out, for example coordinate systems that do not allow the distinction between space-like and time-like vectors, or, more generally, those that extinguish physical differences, are not suitable.[61] In general: The coordinate systems which can be used have to have physical models. Although a coordinate system – as long as it is consistent – cannot be refuted empirically, it still may not be applicable, because it does not have a physical model.

6.3.2 Space: Space can be Understood in a Twofold Way

(A) Space as a coordinate system with a purely mathematical structure, without having physical properties, but with the possibility of receiving a physical interpretation (physical model). Helmholtz[62] found out that the kind of spaces where the distances can be measured with the help of (finite) rigid measuring rods, which are freely movable, are spaces with constant curvature, such that their geometries are either elliptical or hyperbolical or Euclidean. If the rigid measuring rods which are freely movable are infinitesimal, then the spaces are Riemannian, having a Riemannian geometry (Cf. Sect. 4.2).

(B) Space as physical space is a physical reference frame or a physical model of a space coordinate system.[63] A physical model of a space coordinate system, or a reference frame, has to have a physical interpretation of the primitives; that means that it has to have a physical *length standard* or a *unit distance*. Such a unit can consist of the two endpoints of a measuring rod or of two successive maxima of a light wave or of something similar.

---

[60] Schrödinger (1954, STS). See also Möller (1952, ThR).

[61] For example, a purely "prior geometry" that is fixed a priori, i.e. that cannot be changed by changing the distribution of gravitating forces (like the one Nordstrøm has constructed in 1913), could not be suitable for general relativity. For details see Misner, Thorne, Wheeler (1973, Grav) p. 429ff.

[62] Helmholtz (1896, TGZ) p. 1. Recall Sect. 4.2

[63] cf. The distinction of Torretti between physically possible spacetimes and coordinate representations (1983, RaG) p. 302f. For the question of "visually possible space" see Suppes (2002, RIS), Chap. 6.

Although it is conventional which standard or unit distance we choose, it is not only conventional that there has to be a unit, and also the physical properties of such a unit (for example that it is not absolutely rigid, etc.) are not conventional. Observe that physical space is the space of this existing universe or of a part of it. It is not only a container independent of the physical bodies in the universe.

According to Einstein's field equations, our universe is a spacetime with a Riemannian geometry. It has a metric determined by the distribution of the matter and by initial and boundary conditions.

6.3.3 Time: Time can be Understood in a Twofold Way

(A) Time as described by a chronology with a purely logical or mathematical structure, without having physical properties, but with the possibility of receiving a physical interpretation. A mathematical or logical property of a chronology is, for example, that the temporal ordering is a partial ordering (in tense logics also called comparability): x is earlier than y or y is earlier than x or x is simultaneous to y. Moreover, since we want to allow infinitesimal increasing (or decomposing), we assume continuous additivity for a pointwise producing of a line which represents the direction of increasing values on the t (time) – axis.[64] Proposals for chronologies with properties which are rather independent of a physical interpretation are the so-called *tense logics* (logics of time).[65] In such systems, the following properties are usually required by their basic axioms: transitivity, asymmetry, comparability, and density. A relational time theory which abstracts from reference frames (i.e. it is non-local) and can be interpreted as a spacetime structure for Euclidean continuum mechanics was proposed by Noll.[66] In this theory, the first part, with the axioms of time-lapse, time–distance and the definition of Euclidean metric, is a chronology, whereas the second part, which introduces bodies, forces and dynamical processes, is an application to physics (a physical model).

Before we proceed to physical time, three misconceptions connected to the general, chronological concept of time should be mentioned:

(i) It is not correct to say that the time order is reducible to causal order. Although from "event e causes event e' it follows that e is either earlier than or simultaneous with e', the converse is not true.[67]

[64] Note that this has nothing to do with the so-called "arrow of time" nor with time-reversibility or irreversibility. For these notions see Chap. 7.

[65] Prior in his Prior (1957, TMd) was one of the first for such proposals. See also his Prior (1967, PPF). For an overview see Burgess (1984, BTL). cf. further v. Benthem (1991, LgT).

[66] Noll (1967, STS). A similar one was proposed by Bunge (1967, FPh) p. 93ff.

[67] This presupposes, of course, that causal propagation cannot be faster than the speed of light. This, however, is not the place to enter into a discussion of

(ii) It is not correct to say that the law of entropy defines the positive direction of time, because: (a) this law presupposes the concept of time since it states that entropy increases in time, and (b) the concept of time occurs also in theories describing reversible processes and thus must be independent of the law of entropy.

(iii) It is not correct to say that measurements (on the QM level) define the arrow of time, since they produce an irreversible change in the object measured. QM also presupposes the concept of time, and until some years ago we would have added: and further because the laws of QM are time reversible. This, however, cannot be presupposed generally any more, since time reversibility on the micro-level now seems to be violated: not only indirectly via CPT but also directly (see Chap. 7).

(B) Time can be understood as *physical time,* i.e. the time which is shown by a standard clock, or as the time of our universe; of course, without assuming that the time of the standard clock is the time of the universe, or even that there is a unique “time of the universe”. Not every chronology which is logically or mathematically possible is applicable to physical systems (or globally to this universe); only those which have physical models are suitable for physical time. Thus, a chronology which would preclude the relativisation to physical reference systems could not be applied to physics. A physical model of a chronology has to have, further, a physical interpretation of the primitives; that is, it has to define a physical time standard or time unit (or unit interval). This can be done with the help of physical processes which have a certain regularity. There are two types of such processes: periodical and monotone processes. A periodic process is a process where the state of the system repeats itself after a finite period of time and continues to do so in the absence of external disturbing forces. Examples are the spherical pendulum, the planetary movement around the sun, or the daily earth rotation (which was already proposed by Aristotle as a time unit (recall 6.2.2).

A monotone process is a process where the states of the system do not repeat themselves. An example is the uniform (with constant velocity) movement of a body on a straight line. Now, as Mach and Poincaré pointed out, no time standard is determined at any point in space just by empirical reasons. In this sense, the choice of the respective unit is conventional. A time unit can be defined, with the help of a monotone process, by equal distances, since equal distances correspond to equal intervals of time. This presupposes, of course, that we already have the concept of physical space with a length standard or unit distance available.

As for physical time, a further step is necessary: For comparison, one has to have a definition of simultaneity and an operational device for synchronisation

“superluminality”. For a discussion of these issues see: Mittelstaedt, Nimtz (eds.) (1998, SLV). For causality see Chap. 9 below.

of physical clocks. Such a device was originally given by Einstein.[68] Two events on place A and B are happening at the same time if their emitted light rays meet[69] in half of the distance between A and B. Or: two events e and e' (on place A and B) are happening at the same time if they are produced by two light rays coming from a light source which is in the middle of the distance between A and B. Or, in Einstein's description: Two clocks at rest in A and B are synchronised if $(t_B - t_A) = (t'_A - t_B)$. Here, the clock in A shows time t_A when the light ray starts in A, and the clock in B shows time t_B when the light ray arrives in B and is reflected towards A. The clock in A shows time t'_A when the light ray finally arrives back in A. This synchronisation device is called "Einstein synchronisation".

There is, of course, a presupposition involved in this definition of simultaneity: It is assumed that the propagation of light is isotropic (i.e. having the same properties in every direction). It is easily seen that, if the light would travel faster from A to B than back from B to A, the above kind of synchronisation wouldn't work.[70]

6.4 Is Every Law of Nature Spacetime Invariant?

Proposed Answer: The answer to this original question has to be given for different levels. These levels are: Invariance under internal spacetime transformations (6.4.3); invariance under inertial transformations I (Galilean transformations) (6.4.4); invariance under inertial transformations II (Lorentz transformations, special relativity) (6.4.5); and invariance under arbitrary spacetime transformation (general relativity) (6.4.6).

6.4.1 General Remarks on Invariance

As is clear from 5.1 and 5.3.1, invariance or symmetry can be applied to both physical systems and laws (of nature) describing such physical systems. Here, we shall concentrate on the invariance of laws[71] and on those parameters (magnitudes) occurring in the laws.

Concerning the invariance of laws of nature, we can distinguish different degrees of generality:

[68] Einstein (1905, EBK) p. 894.

[69] In order to point out the relation to a reference system we may say: "meet at an observer".

[70] For the possibility of different definitions of synchronisation and consequently for simultaneity see below, Sect. 6.4.5, for a detailed discussion see Mittelstaedt (1989, ZBP) p. 41ff.

[71] It will be understandable that consequences of laws are not discussed here. They may be much less general, and especially if they are inferred with the help of initial or boundary conditions like predictions (for example as solutions of an equation of motion), they are much too restricted to be invariant at all in a similar way as the laws.

(1) In a very general sense, a law is invariant if it does not change when other things change. This is the general understanding of a law already described in 5.3.1(1).
(2) In one specific sense, a law (of nature) is invariant if it does not change when space (place) or time change. This is the oldest and – at least for our understanding of what a law is – the most important invariance of laws of nature. Its traditional and present importance has already been discussed with an example in 5.3.1 above. Its deeper features will be discussed below.
(3) In a sense more general than (2) but less general than (1), a law of nature is invariant if it does not change under a change of (physical) reference frames (or under a transformation of reference frames) (cf. 5.3.2(2)). This is the sense which is used nowadays to formulate most invariance principles.

It will be understood that reference systems of spacetime have an upper limit for their symmetries: There are at most seven parameters for internal transformation (three for translation in space, three for orientation in space, one for translation in time) and three for inertial movement. That is, the symmetry groups of Galilean – and of Lorentz – transformations have at most ten parameters each.

(4) There are also additional kinds of invariance:
  (a) Dynamical symmetries
  In addition to the spacetime symmetries mentioned, some physical systems admit another kind of symmetries, the so-called dynamical or internal symmetries. The most interesting example is the "Kepler problem", i.e. the motion of a particle in a central field with a $1/r$-potential. Here, the Hamiltonian is invariant with respect to a three-parameter infinitesimal canonical transformation of the phase space, whose generator is given by the Lenz–Runge vector Λ_i. Accordingly, this vector is a conserved quantity of the Kepler problem, which is, however, not induced by a spacetime symmetry in the Minkowski spacetime as rotation or translation, say.
  (b) Form invariance
  There are other changes of the coordinate system which change the form of a law. For example, if rectangular coordinates are changed into spherical coordinates, then the law formulated with its help changes its form, though it does not change its validity or its invariance properties. From this it is clear that *invariance of laws* differs from *form invariance of laws*. The same law may be expressed in different forms, and one reason for such a difference may be the coordinate system (cf. 6.4.2).

Concerning parameters (and magnitudes) occurring in laws, a physical parameter (magnitude) is invariant if it does not change its value under a change from one (physical) reference frame to another.

6.4.2 Spacetime Invariance is Concerned with Real Continuous Movements of the Reference Frame

This can be seen as follows: Concerning movements, we have to distinguish between real movements and virtual movements. A physical reference frame can only make real movements like translations and rotations. A coordinate system, however, can also make virtual movements like mirror reflections. Since it is a mathematical theorem that every position (orientation) and movement of an object in space can be produced by a suitable sequence of translation, rotation, and mirror reflection, it is a natural question whether laws of nature can be invariant under any (also virtual) change of orientation and movement of the coordinate system. The answer to this question is: No. The reason is that laws of nature are not in general invariant under mirror reflection. This is known since Yang and Lee discovered that parity is violated in nature. Now, since physical reference frames cannot transform into their mirror image, parity has been listed under "Discrete Symmetries", a different kind of symmetry compared to "Continuous Spacetime Symmetries" (recall 5.3.2 (2) and 5.3.2 (3)). Therefore, spacetime invariance is invariance of the laws under continuous transformations. Groups of such transformations are mathematically describable by Lie groups (of various parameters).

6.4.3 Invariance Under Internal Transformations

Concerning transformations, we first have to distinguish between two types of transformations. According to the first type of transformations, only the comoving spacetime coordinates within a given frame of reference are changed by transformations. This means that here we merely consider translations of the spacetime coordinates and rotations in space. According to the second type of transformations, the velocity field which characterises the reference frame is also changed by convenient transformations.

The first type of transformation will be called internal transformation. The reason is that such a transformation is nothing but a replacement, within the system, of the respective coordinates by new ones which have a constant distance (or constant angle) from the original. Thus, the space coordinate x is replaced by $x' = x + a$ (as for the other two); the time coordinate t is replaced by $t' = t + b$, and the rotational (orientation) angle α is replaced by $\alpha' = \alpha + \beta$ (as for the other two).

What happens with such transformations is expressed very vividly by Stephen Weinberg. The first three transformations in this passage refer to the first type of transformation, the last one to the second type:

> "The paradigm for symmetries of nature is of course the groups of symmetries of space and time. These are symmetries that tell you that the laws of nature don't care about how you orient your laboratory,

or where you locate your laboratory, or how you set your clocks or how fast your laboratory is moving."[72]

As has been mentioned earlier (see Sect. 5.3.2 (2)), the invariance of laws of nature under these three transformations leads to seven conservation principles: three for momentum, three for angular momentum, and one for energy.

Observe, however, that this first type of transformations, i.e. the internal transformations, abstracts from movement of the reference frame. In other words, they do not take into account a movement of the reference frame or laboratory. Therefore, they are also abstract from fast movements (where velocity v is close to the order of magnitude of c) or from accelerations and consequently from (strong) forces and (strong) fields. Although these transformations abstract from forces and fields, it would be a misunderstanding to say that they implicitly claim (or even make the assumption) that there are no forces or fields. Such a claim or statement is not a consequence nor implied by that first type of transformation.[73]

From the above considerations a first-level answer follows to our original question "Is every law of nature spacetime invariant?" It reads: Granted that we abstract from high velocities and from strong fields and forces, every genuine law of nature is spacetime invariant, in the sense of continuous translation with respect to internal transformations.

This first type of invariance (under internal transformations) forms a subgroup of the so-called Galilean group: What is relevant here is only translation in space (G_T), translation in time (G_t), and rotation in space (O_3), i.e. $G_T \times (O_3 \times G_t)$. The subgroups space translation and rotational translation form the Euclidean group or the geometrical transformations in Euclidean space; whereas for the full Galilean group G we have to add the change of inertial systems (G_o):

$$G = (O_3 \times G_t) \times (G_T \times G_o) \, .$$

According to this first type of internal invariance, we can say – always abstracting from forces and fields – that space is isotropic, and time is homogeneous.

6.4.4 Invariance Under Inertial Movement

A reference system can move in two ways: As an inertial system, without acceleration and without inertial forces, and as a system with acceleration

[72] Weinberg (1987, TFL) p. 73.

[73] Observe that the *concept* of transformation and invariance with respect to the first type is independent of historical facts. We mean, for example, historical facts in the sense that Galileo might have thought that the inertial movement of the ship (in his example in his (DWS), Second Day) can have arbitrary velocity – relative to a ship at rest (or slowly moving) – without any consequence on the measuring rods of his laboratory on the ship. Here we might speak of some false assumptions, which have been corrected by special relativity.

and inertial forces respectively. Here, we restrict our considerations to the motion of inertial systems. Generally, an inertial system moves with constant velocity relative to any other inertial system. Note that any movement of an inertial system A is an inertial movement relative to another system B, even if we assume that B is at rest relative to A. Since there is no absolute space, there is no absolute reference point available relative to which we could judge that system A moves inertially, i.e. with uniform velocity on a straight line. Therefore, the class of all inertial systems can be taken as a realisable substitute for the metaphysical concept of absolute space.[74]

6.4.5 Invariance Under Inertial Movement I: Galilean Movement

Inertial systems of reference are characterised by the requirement that force-free point-like particles (mass points) move along straight lines. It follows from this condition that two inertial systems $I$and I' move relatively to each other with a constant velocity v. Obviously, it is assumed in this argument that there are no global fields like gravitation in particular fields which could destroy the isotropy and homogeneity of space. Such fields, though they do not destroy the invariance of laws, violate the necessary conditions (or presuppositions) such as the isotropy of space, under which inertial systems can exist at all. On the basis of this result, we can give a second-level answer to the original question: Every genuine law of nature is invariant under mutual transformations of inertial frames of reference with a constant relative velocity v. This is a kind of "relativity principle" which can also be expressed in the following way: *Laws of nature do not distinguish between inertial systems moving relative to each other with a constant velocity v.* Or: *Inertial systems moving relative to each other with a constant velocity v are equivalent with respect to laws of nature.*

This statement should be further specified and illustrated by an example: In one spatial coordinate the equation of motion in classical mechanics reads $m(d^2x/dt^2) = K_x$. If we assume – in accordance with Newton – that there is a universal time parameter t which is relevant for all inertial systems, then the transformation from system $I(x, t)$ to the system I' (x', t') is given by the Galileo transformation $x' = x - vt$, $t' = t$. Obviously, the equation of motion mentioned is invariant against this transformation. However, if we do not assume – in accordance with Einstein – that there is a universal time t, then the Galileo transformations must be replaced by a Lorentz transformation, and Newton's above-mentioned equation is no longer invariant. Hence, we must replace the classical equation of motion by the Lorentz invariant relativistic one, which leads to results which are unknown in Newtonian mechanics. For example: The inertial mass of a moving body depends on its velocity v according to the relation $m(v) = m(0)/\sqrt{(1 - v^2/c^2)}$, where c is the velocity of light. It should, however, be emphasised that the differences between the classical and the relativistic case are almost negligible if we restrict our considerations

[74] cf. Mittelstaedt (1989, KLM) p. 44.

to velocities $v \ll c$ that are very small, compared with the velocity of light in vacuum. In this case, the well-known consequences of Newton's equation of motion, the velocity independence of mass, length, and time periods, are almost in accordance with experimental results.

In addition, it should be noted that "Galilean invariance" and "Galilean transformation" are not always described in the same way. A reason for this is that they can also be understood historically. From the historical point of view, in classical mechanics there is no restriction on the concept of time and consequently on the velocity of the moving inertial systems; i.e. v could not only come close to c but also exceed c.[75] Even if the latter is at least controversial, if not impossible, according to our knowledge today, a velocity v that approximates c has the well-known serious consequences (of which an important example is given above).

Summing up, it should be noticed that the Galilean invariance made three important hidden assumptions:

(i) The time scale is the same in all inertial systems.
(ii) Simultaneity is the same in all inertial systems.
(iii) The spatial distance of two simultaneous events is the same in all inertial systems.

6.4.6 Invariance Under Inertial Movement II: Special Relativity

Inertial systems (frames) move with $v \leq c$ relative to each other or undergo internal transformations (6.4.3), thereby abstracting from forces or fields. In his special theory of relativity,[76] Einstein's problem was a question of the invariance of physical laws: Is it possible to formulate both the laws of classical mechanics (Newton's theory) and the laws of classical electromagnetism (Maxwell's theory) in such a way that both laws are invariant under transformations of inertial systems?

> "The same laws of electrodynamics and optics are valid for all frames of reference for which the equations of mechanics hold good ... We will raise this conjecture (the purport of which will hereafter be called the 'Principle of Relativity') to the status of a postulate and also introduce another postulate which is only apparently irreconcilable with the former, namely that light is always propagated in empty space with a definite velocity c which is independent of the state of motion of the emitting body. These two postulates suffice for the attainment of a simple and consistent theory of the electrodynamics of moving bodies based on Maxwell's theory for stationary bodies."[77]

[75] See the passage of Galileo (DWS) in Sect. 5.4.2
[76] Einstein (1905, EBK) Engl. Transl. p. 37f.
[77] Ibid.

Since Maxwell's equations are not Galileo invariant, there were three main options: (i) Assume an ether which is the designated inertial frame for Maxwell's theory, whereas the laws of classical mechanics have their Galileo invariance. (ii) Try to adapt Maxwell's equations to Galileo invariance by correcting them, so that both laws are Galileo invariant.[78] (iii) Invent a new principle (or kind) of invariance which is obeyed by Maxwell's equations and adapts classical mechanics (Newton's theory) by correcting it.

Einstein had chosen (iii) even before experiments had confirmed his choice. That he rejected (ii) is certainly connected with the fact that (ii) is incompatible with his strong belief – supported by his knowledge of electrodynamics – that in general the velocity of light is constant, and its propagation is independent of the emitting source.

(i) Was rejected by the Michelson–Morley experiment and other experiments (Kennedy–Thorndike experiment and stellar aberration), which refuted alternatives (Lorentz–Fitzgerald contraction, ether-drag hypothesis) for saving an ether.
(ii) Was refuted by experiments which showed the independence of the velocity of the source (the first experiments were done by de Sitter in 1913 and were later confirmed by experiments with the help of particle accelerators). Einstein's new principle of invariance was the so-called Lorentz invariance, and his correction of the laws of classical mechanics led to the Lorentz invariant mechanics.

Historically, it is worth mentioning that Lagrange had already corrected Newton's equations, in order to show how Newton's absolute motion could be deduced from observed relative motions. He formulated Newton's equations (for three gravitationally interacting bodies) in such a way that they contained, besides the time coordinate, three only relative space coordinates (determining the sides of a triangle in the corners of which are the three bodies).[79]

(a) Lorentz invariance
Lorentz invariance is invariance under transformation of inertial systems, without assuming that the time scale (time measurement) is the same in all inertial systems. This invalidates three important hidden assumptions of Galilean invariance: (α) There is an absolute time in all inertial systems, i.e. the time scale (time measurement) is the same in all inertial systems. (β) Simultaneity is the same in all inertial systems; and as a consequence (γ) the spatial distance of two simultaneous events is the same in all inertial systems. That means that according to Galilean invariance time measurements, simultaneity, and spatial distance are invariant magnitudes under transformations of inertial systems. And these three invariances were invalidated by Einstein's theory of special relativity.

[78] This option was investigated by Stachel and Jammer (1979, MWA).
[79] Lagrange (1772, EPT). For a detailed account see Dziobek (1888, MTP). cf. Barbour (2001, GCB) p. 200.

(b) Lorentz transformation
The Lorentz transformation transformes the coordinates x, y, z, t of the inertial system I into those x', y', z', t' of the inertial system I', which moves with velocity v with respect to I:

$$x' = k(x - vt) \qquad y' = y \qquad z' = z$$

$$t' = k(t - \tfrac{vx}{c^2}) \qquad \text{where} \qquad k = \frac{1}{\sqrt{1 - v^2/c^2}} .$$

As one can easily see, the decisive difference between the Galilean transformation and the Lorentz transformation is that in the former simply $t' = t$, whereas in the latter $t' = k(t - vx/c^2)$.

Consequently, (for velocities approaching c) there is contraction of length (by the factor $1/k$), dilatation of time (by the factor k), and increasing mass (by the factor k).

Meanwhile, all three consequences have been experimentally confirmed: The increase of mass by experiments with nuclear particles in accelerators; the dilatation of time by comparing the mean life for decay of muons in flight with velocity close to c with muons brought to rest in an absorbing block. Later, the effect could even be proved with atomic clocks in an aeroplane.[80] The contraction of length in the direction of movement of an object can only be observed as an object which appears to be turned (and thus shorter).

(c) Invariance of magnitudes
The most important invariant magnitude is, of course, c, the velocity of light. No change in the inertial reference frame or in the velocity of the light source can change the speed of light; it is invariant under all changes of inertial systems. In other words: The propagation of light is isotropic with respect to all inertial systems. This was first confirmed by Michelson's experiment (1882–1887). Another invariant magnitude is electric charge. This basic magnitude in Maxwell's equations is Lorentz invariant and does not need any correction.

Although mass, length, time, and simultaneity are not invariant and have to be corrected as shown above, there are combined magnitudes of the above which are invariant. This can be seen by the following consideration:

In an inertial system I with coordinates (x, t), we consider two events $E_1(x_1, t_1)$ and $E_2(x_2, t_2)$, where we restrict our considerations to one space coordinate. Neither the spatial distance $\Delta x = x_2 - x_1$ nor the time difference $\Delta t = t_2 - t_1$ of these two events are Lorentz invariant, since in another inertial system $I'(v)$ moving with velocity v the transformed quantities read $\Delta x' = k(v)(\Delta x - \Delta vt)$ and

$$\Delta t' = k(v)(\Delta t - v\Delta x/c^2) .$$

[80] This was first done by Hafele and Keating in 1971.

The “four-dimensional distance”

$$(\Delta s)^2 = c^2(\Delta t)^2 - (\Delta x)^2 \ ,$$

however, is Lorentz invariant, since from the transformation formulas for Δx and Δt it follows:

$$c^2(\Delta t')^2 - (\Delta x')^2 = c^2(\Delta t)^2 - (\Delta x)^2 \ .$$

By means of the spacetime distance s, other Lorentz invariant quantities can be constructed.

(d) Invariance of laws
We can now give a third-level answer to our original question: *All laws of electrodynamics (electromagnetism), i.e. of Maxwell's theory, and all laws of classical mechanics which are corrected in the sense of special relativity are invariant under transformations of inertial systems moving with* $v \leq c$. Since inertial systems move along spacetime coordinates (physical objects of the inertial system move along a world line), invariance under transformations of inertial systems (moving with $v \leq c$) also means spacetime invariance. Another description of the same fact is that all the laws of nature mentioned above are Lorentz invariant. Again, another way to express these facts is that there is no designated or “true” inertial system from which time, place (space), mass, or velocity is universally measured (for all other reference frames). On the contrary, every inertial reference frame is – objectively – of equal rank. This can also be expressed by the following three points: There are no privileged points in space and time, there are no privileged directions in space, and there are no privileged inertial systems. These formulations of the invariance of laws of nature are at the same time alternative formulations of the principle of special relativity (recall Einstein's citation in 6.4.6). If we replace the third point by “there are no privileged frames of reference at all” we pass to one of the most important guiding ideas of general relativity. As is clear from the discussion in 5.3.1 (3), such principles can be expressed as meta-laws (either in descriptive or in normative form) or as laws about physical systems.

6.4.7 Invariance Under Arbitrary Spacetime Transformations: General Relativity

As in the case of special relativity, Einstein's main idea for developing his theory of general relativity (GR) was again invariance of laws of nature. Invariance of laws in the most general sense transcends Lorentz invariance in a threefold way: first, in the sense that it drops the restriction to inertial reference frames (a), second, in the sense that it drops the restriction to straight Galilean coordinates (b), and third, in the sense that it extends to gravitation (6.4.7.1), which was left out of consideration by special relativity.

(a) To drop the restriction to inertial reference frames is well justified, for many reasons. Four important ones are as follows:
   (i) Because rotational movements, like the movement of the planets around the sun and their self rotation, are not inertial motions, since they are not free from inertial forces (in this case centrifugal forces, Coriolis forces, etc.), they must be described by means of accelerated frames of reference.
   (ii) In general, accelerated movements have to be permitted in addition to inertial movements.
   (iii) Mach's reinterpretation of Newton's bucket experiment showed that the motion of the water relative to the bucket does not lead to inertial forces, whereas the motion relative to the starry sky induces centrifugal forces. Mach left the question open how the stars could act on the rotating water, but Einstein – more than twenty years later – assumed that the gravitational field of the distant masses of the universe induces inertial forces. Einstein called this hypothesis "Mach's principle" and considered it as one of the corner stone of the new theory of general relativity (see 6.2.1.6 and 6.2.2.4).
   (iv) The existence of gravitational fields and the impossibility to screen off gravitational fields (of planets, stars and galaxies) precludes the construction of global inertial systems, and thus, strictly speaking, of inertial systems at all.

(b) The restriction to straight Galilean coordinates had to be given up for the same reason. Because there are no inertial systems (at least not globally), the geometry of light rays in a three-dimensional reference frame which moves with acceleration is not Euclidian. In general, the world line of light rays and of physical systems (reference frames) is geodesic, which means as straight as possible. More accurately: these are geodesic lines in the Riemannian spacetime, which is constituted by the local inertial systems. Consequently, in the presence of gravitational fields we must restrict all considerations to *local inertial systems* I (x_k, t), i.e. to reference systems which are inertial only at one place x_k and at one instant t of time. This restriction has far reaching consequences for the geometry of spacetime.

Summing up, the general theory of relativity, which is in fact a general theory of gravitation, can be characterised by the following four basic assumptions:

(1) There is a spacetime metric.
(2) This metric is related to the distribution of matter and energy in spacetime by Einstein's field equations, with the following consequence: The effect of gravitation on matter fields can be described by replacing the flat Minkowski metric of special relativity by a curved Riemannian metric.
(3) Assuming the principle of equivalence between inertial and gravitational mass, the formalisms which describe inertial forces can be applied to gravitational forces as well.

(4) All special relativistic laws (laws of special relativity) are valid in local Lorentz frames of the metric.[81]

6.4.7.1 The Principle of Equivalence

The incorporation of gravitation is essentially connected with dropping the restriction to inertial movement and incorporating accelerated frames of reference. It was one of the important basic ideas in the development of general relativity to see that, with respect to the motion of material bodies, a homogeneous gravitational field is indistinguishable from and can be replaced by a uniformly accelerated frame of reference. Hence, a reference system K, which is at rest in a homogeneous gravitational field, is equivalent to a reference system K' (in a gravitational free space) moving with a uniform acceleration. Since homogeneous gravitational fields do not exist in reality, these arguments must be restricted to local situations: With respect to the motion of a mass point, the influence of a gravitational field can locally and momentarily not be distinguished from the effect of an accelerated frame of reference. Consequently, locally and momentarily a gravitational field can be "transformed away" by a transformation to a local and momentary inertial frame of reference.

This principle is called the *principle of equivalence*.[82] It is usually formulated like this: Physical laws are the same in all "free falling" local inertial reference frames and satisfy in these systems the laws of special relativity. Or in other words: Physical laws are invariant under changes of "free falling" local inertial reference frames. Both formulations, however, are not sufficiently detailed. Therefore, we want to add the following details:

(1) The equivalence is claimed between gravitational effects and acceleration or between effects in "free falling" and "inertial" frames of reference. In the first case, between a reference frame at rest in a gravitational field and a reference frame moving with acceleration, and in the second case, between a free falling reference frame and a reference frame with inertial movement (free from forces). An example for both cases is Einstein's "Gedankenexperiment" with the elevator. With respect to the motion of

[81] cf. Misner, Thorne, Wheeler (1973, Grav) p. 302 and Hawking (1980, TAG) p. 145. For a detailed discussion of basic concepts and axioms see Bunge (1967, FPh) p. 218ff.

[82] In the respective literature, this principle is sometimes called *weak principle of equivalence* in contradistinction to the *strong principle of equivalence*, which is the statement that the inertial mass and the gravitational mass are "equivalent" or better: proportional with a universal factor. The usage of the terms "weak" and "strong" is probably connected with the idea that the "weak" principle is in some way derivable from the "strong" one. We do not use these terms, because they are used in other contexts with another connotation. Such terms may therefore be misleading.

a mass point, the effect of a gravitational field in a given spacetime point cannot be distinguished from an accelerated frame of reference, and a free falling reference frame cannot be distinguished from an inertial reference frame.

(2) A local reference frame relative to which a static gravitational field vanishes is equivalent to an inertial frame (because it is free from forces). And a local reference frame relative to which a static gravitational field emerges is equivalent to a uniform accelerated frame.

(3) It is always possible to make such a transformation that in the reference frame a static homogeneous gravitational field vanishes. And: It is always possible to make such a transformation that in the new reference frame an acceleration of a particle vanishes. Moreover, a non-static and non-homogeneous gravitational field is equivalent only to a local-instantaneous reference frame.[83] On the other hand, a non-static and non-homogeneous gravitational field can vanish locally and, at a fixed point of time, only through a local-instantaneous reference frame. Reference frames of this kind may be realised by free falling reference frames. Although they are locally inertial at every point of time, they have to be newly defined for every point of time. Accordingly, a free falling observer (for example in a satellite) could adjust to continually new inertial systems.

(4) It is usually assumed that the explanation or reason (or the premise) for the principle of equivalence described by (1)–(3) is the equivalence of inertial mass and gravitational mass, or, more accurately, the fact that inertial mass and gravitational mass are proportional with a universal factor. This fact itself does not have a satisfactory explanation so far.

With respect to the role of the principle of equivalence for GR we want to point out three things: First, it was heuristically important for the development of GR; second, this principle leads at most to the Riemannian structure of spacetime but not to the Einstein equations, which describe the relations between the spacetime metric and the sources of the gravitational field; third, the principle of equivalence does not imply the regulating guidance principle of GR, which has been called the principle of general covariance of laws of nature.

6.4.7.2 General Invariance and General Covariance

As a guidance to finding the field equations of the gravitational field, Einstein used a regulative principle which states that the laws of nature should be independent both of any arbitrarily moving reference system and of any coordinate system.[84] In Einstein's words:

[83] cf. Misner, Thorne, Wheeler (1973, Grav), p. 386.

[84] A reconstruction of this rather hypothetical "derivation" of Einstein's field equations in modern terminology can be found in Anderson (1967, PRP) pp. 338–342.

> "The laws of nature must be of such a nature that they are valid with respect to systems of reference in arbitrary motion"
>
> "The general laws of nature are to be expressed by equations which hold in all systems of coordinates, i.e. which are covariant with respect to any arbitrary substitutions (generally covariant)."[85]

The two quotations show that Einstein distinguished between two aspects of independence of the validity of laws. The first aspect (here called A) focuses on physical reference systems under arbitrary motion; i.e. laws of nature remain valid under arbitrary motion of physical reference systems. The second aspect (here called B) focuses on mathematical coordinate systems under arbitrary transformation;[86] i.e. laws of nature remain valid under an arbitrary transformation of mathematical coordinate systems.

The validity of laws of nature, in the sense of A, is sometimes called the *principle of general invariance,* [87] but usually referred to as the *principle of general covariance*. Also, the validity of laws of nature, in the sense of B, is called the *principle of general covariance*. There is no generally accepted terminology with respect to the expressions *invariant* and *covariant*, and there is still a good deal of confusion concerning what content Einstein implied by the principle of covariance.[88]

Concerning the aspects A and B, we shall deal with the following questions: (1) What is the connection with the theory of general relativity? (2) How can the *relativity principle* be expressed? (3) Are both aspects A and B realised

---

[85] Einstein (1916, GAR), p. 772 and 775f. Engl. Transl. p. 113 and 117. The German text reads: "Die Gesetze der Physik müssen so beschaffen sein, daß sie in bezug auf beliebig bewegte Bezugssysteme gelten. Wir gelangen also auf diesem Wege zu einer Erweiterung des Relativitätspostulates." "Die allgemeinen Naturgesetze sind durch Gleichungen auszudrücken, die für alle Koordinatensysteme gelten, d.h. die beliebigen Substitutionen gegenüber kovariant (allgemein kovariant) sind."

[86] We do not deny that Einstein sometimes did not distinguish both aspects, or that he sometimes mentions only one aspect and later speaks of point coincidences etc. It also appears from Einstein's text that immediately after the second quotation he claims it implies the first (the principle of relativity): "It is clear that a physical theory which satisfies this postulate will also be suitable for the general postulate of relativity." That this is not the case will be shown in (5) below. cf. also Friedman (1983, FST) p. 207f., and Norton (1989, CCE) and Norton (1993, GCF). For further historical discussions see Howard (1999, PCP) and Rynasiewicz (1999, KAC). For our purpose, to make clear the two aspects it is sufficient that Einstein at least formulated initially this difference.

[87] cf. Anderson (1967, PRP) p. 338.

[88] As Thirring (1979, CMP) p. 166 says: "At the time of the birth of gravitation theory, the requirement of general covariance provided some relief from labor pains, but later on, it was more often a source of confusion". For a detailed review of the different interpretations of *general covariance*, which have been proposed since the twenties, see Norton (1993, GCF).

in and bound to general relativity? (4) Can both aspects be reduced to the concepts of *active* and *passive* transformations? (5) Are both aspects A and B equivalent?

(1) Aspect A refers to the equivalence of reference systems in arbitrary motion and is an expression of the principle of relativity. However, since in spite of a misleading terminology the theory of general relativity does not fulfil the principle of relativity, aspect A does not apply to and is not realised in the theory of general relativity.

  Aspect B refers to the equivalence of arbitrary coordinate systems with respect to the laws of nature. This equivalence of systems of coordinates can be achieved by reformulating the laws of nature in terms of tensors, the calculation rules of which guarantee the covariance of the laws considered.[89] Since the theory of general relativity is a covariant formulation of the theory of gravitation, aspect B is fully realised in general relativity.

  Of the three versions of the principle of general covariance, in the sense of B, distinguished by Carmeli, the second seems, therefore, to be the most uncontroversial if applied to GR:[90] (i) All coordinate systems are equally good for stating the laws of physics. (ii) The equations that describe the laws of physics should have tensorial form and be expressed in a four-dimensional Riemannian spacetime. (iii) The equations describing the laws of physics should have the same form in all coordinate systems. This is so because version (i) depends at least partially on pragmatic goals, and version (iii) is not acceptable, since covariance (invariance) of laws differs from form invariance of laws, as has been pointed out already in 6.4.1 (4b).[91]

  On the other hand, the connection of aspect B to GR is not special or unique. This has been shown first by Cartan (1923) and Friedrichs (1927), who gave coordinate free formulations of Newton's theory of gravitation.[92]

(2) There is the question of how the principle of relativity should be expressed. Should it be expressed by aspect A or by aspect B or by both?[93] Let us

---

[89] This was Einstein's key for the tensor formulation of his field equations. cf. Einstein (1916, GAR) p. 780, Engl. Transl. p. 121.

[90] Carmeli (1982, CFG) sections 1.4 and 1.5. cf. also the discussion in Norton (1993, GCF) p. 817.

[91] That and how form preservation restricts too much was shown also by Post (1967, PNG).

[92] See Misner, Thorne, Wheeler (1973, Grav) Chap. 12 for more about "Newtonian Gravity in the Language of Curved Spacetime", and Penrose (2001, GRQ) p. 298ff. cf. also the formulation of the Kepler problems in terms of Lagrangian mechanics with generalised coordinates.

[93] The formulation by Einstein of the principle of relativity for SR in Einstein (1905, EBK) was expressed with aspect A; the "need for an extension of the postulate of relativity" was formulated in Einstein (1916, GAR) Sect. 2 with aspect A and in Sect. 3 under the title "... Covariance for the Equations..." with aspect B. Aspect B is also used already in Einstein (1911, ESA). The question of expressing such

briefly recall the meaning of the principle of relativity as it is referred to in many present-day textbooks.[94] Consider two local frames of reference R and R′ and two processes P and P′ which are prepared and performed in the same way in the two reference systems R and R′, respectively. The principle of relativity, then, claims that the two processes P and P′ are equivalent and lead to the same observable results. This means that also the laws of nature which govern the processes P and P′ can be formulated such that they have the same form in R and in R′. In other words, it is not possible to distinguish the reference frames R and R′ by local processes P and P′ and by the corresponding laws of nature.[95] It is well known that this principle of relativity is valid for two inertial systems I and I′. Here, however, it is postulated for arbitrary local frames of reference.

The principle of relativity can formally be expressed by the requirement that the laws of nature must be covariant under coordinate transformations between two local frames of reference. This formulation corresponds to aspect B and is also called *general covariance* by some authors. Expressed as aspect A, the principle of relativity claims that a physical process and the laws which govern this process are invariant under changes of the observer's reference system. This invariance is sometimes called *general invariance*. [96]

(3) A further question is whether the ideas expressed by aspect A and aspect B are realised in or restricted to the theory of general relativity (GR), i.e. whether they are restricted to the present relativistic theory of gravitation.[97] The answer to this question is negative, for the following reason. Although these ideas were guiding principles in the development of GR, they transcend GR in that they are supposed to hold not only for a most general theory of gravitation, but universally for every correct theory; i.e. the idea behind is that the laws of nature are the same in the whole universe. And, consequently, if the laws satisfy aspect A or aspect B, then no particular system is preferred or designated by these laws: The laws do not select or designate any particular (physical) reference frame or any particular (mathematical) coordinate system. As mentioned already in (1) above, it has been shown in an important particular instance that aspect B is not restricted to GR by providing coordinate free formulations of Newton's theory of gravitation.

(4) Can both aspects A and B be reduced to the concepts of *active* and *passive* transformations? Generally, an *active* transformation changes one vector

---

principles in a descriptive or normative form and in a meta- or object-linguistic form has been discussed in detail in 5.3.1 (3).

[94] cf. Sexl, Urbantke (1992, RGT) p. 55ff.

[95] Recall Galileo's thought experiment discussed in Sect. 5.4.2.

[96] cf. Anderson (1967, PRP) p. 338, and Schmutzer (1996, RTA) p. 115f.

[97] Here, we refer to the standard version of the theory of gravitation which is subject to many textbooks, e.g. Rindler (1977, ESR). There are, however, generalisations and modifications of this theory which will not be considered here.

into another, while leaving unchanged the underlying reference frame. By contrast, a *passive* transformation leaves all vectors unchanged but alters the reference frame. For measurable physical objects which are expressed by vectors, tensors etc., active and passive transformations are in some sense equivalent, expressing a duality in describing the same object.

(5) Are both aspects A and B equivalent in physics?
This question can be subdivided into the following two subquestions:

Subquestion 1: Is every change of a reference system representable by a corresponding transformation of a coordinate system?

Subquestion 2: Is every transformation of a coordinate system representable by a corresponding change of the reference system?

It will be shown, subsequently, that there is no general equivalence between aspects A and B. This will be done in three steps. Firstly (a), two general remarks concerning aspects A and B will be made. Secondly (b), it will be shown that there is no equivalence if the aspects A and B are applied outside their usual area of application to non-continuous (discrete) transformations. Thirdly (c), it will be shown that within the usual area of application, there is only equivalence on some restricted levels.

(a) General remarks
To every reference system many different spacetime coordinate systems can be associated.[98]

These different coordinate systems are, however, not completely arbitrary. They are restricted by methodological constraints which are suitable for their area of application; such constraints are, for example, that the direction of time is not reversed, or that the topology of causality is preserved.[99] Regardless of this restriction, it is obvious that each reference system can be associated with several coordinate systems. Starting from coordinates $x^{\mu}(\mu = 0, 1, 2, 3)$, any coordinate system $x^{\mu\prime}$ given by $x^{k\prime} = f^k(x^1, x^2, x^3), x^{0\prime} = f^0(x^0, x^1, x^2, x^3)$ describes the same reference system, using only different space coordinates and different clock synchronisations. For example, the transformation from spatial Cartesian coordinates to polar coordinates is not representable by a corresponding change of reference systems. Also, the change from Einstein synchronisation ($\varepsilon = 1/2$) in an inertial system to another clock synchronisation, $\varepsilon = 1/3$), say, cannot be achieved by changing the reference system. Conversely, a spacetime coordinate system determines a reference system only under the condition that it is a *comoving* system such that the material elements of

[98] Remember that according to our terminology introduced in 6.3, a coordinate system is a mathematical structure (homomorphism), whereas a reference system has a material basis which is equipped with rods and clocks.

[99] Such methodological constraints have been elaborated by Heintzmann, Mittelstaedt (1968, PGB). See also: Time Reversal, below Sect. 7.2.3.4.3.

the reference system have temporally constant spatial coordinate values. Thus, subquestion 2 is answered negatively, and therefore, there is no equivalence between aspects A and B.

(b) Aspects A and B applied to non-continuous transformations
It can be shown quite clearly that in the following two interesting cases of non-continuous (discrete) transformations, subquestion 1 is answered positively, but subquestion 2 is answered negatively:

Time reversal: Time reversal can be easily represented by a transformation of the coordinate system, by transforming signs "t" into signs "$-t$". This does not correspond, however, to a real change of the reference system (or laboratory), which cannot move backwards in time; the reference system can cross the $t =$ constant line only in the direction of increasing t. Observe, further, that time reversal is not generally satisfied in weak interaction (cf. Sect. 7.2.3.4.3).

Parity: Mirror image reflection can be represented by a transformation of the coordinate system, but this does not correspond to a change of the reference system (or laboratory), which cannot switch into its mirror image with its rods and clocks; although one could construct new measuring rods and clocks which are the mirror images of the old ones. Observe, further, that parity is not generally satisfied at the microlevel (in weak interaction).

(c) Aspects A and B applied to continuous transformations

Internal transformations: These are transformations which correspond to the first seven parameters of the Galileo group, i.e. translation and orientation in space, and translation of time without movement (cf. 6.4.3 above). The most general internal transformation is given by

$$x^k \to x^{k\prime} = f^k(x^1, x^2, x^3) \quad (k = 1, 2, 3) \; ; \qquad x^0 = f^0(x^0, x^1, x^2, x^3)$$

In this case, subquestion 1 is trivially satisfied, but subquestion 2 has to be answered negatively.

Galilean transformation: Concerning Galilean transformation, there is only an equivalence between aspects A and B if the (hidden) assumption of the universal temporal metric (which implies also arbitrary velocities) is not physically used. Under this restriction, laws of nature do not distinguish among reference systems which correspond to Galilean coordinate transformations such that aspects A and B are equivalent. On the other hand, there is no such equivalence in general if the assumption of a universal temporal metric, which allows arbitrary velocity and arbitrary transportation of measuring rods and clocks, is not restricted.

Special relativity (SR): Concerning Lorentz transformation, there is an equivalence between aspect A (invariance of laws under change of reference systems) and aspect B (covariance of laws under

transformations of coordinate systems), provided the reference systems considered are inertial systems. Hence, all inertial systems are equivalent. And, consequently, aspects A and B are equivalent, in this restricted sense, since changes of inertial frames of reference and transformations of inertial coordinate systems are mutually representable.

On the other hand, in the presence of gravitational fields there are only *local* systems of inertia (local geodesic systems of reference). Finitely, extended reference systems cannot be systems of inertia. Hence, there is no equivalence between aspects A and B such that subquestion 2 has to be answered negatively.

The non-equivalence between aspects A and B is evident from the following further point: The principle of relativity that all laws are the same under change of inertial reference systems – which corresponds to aspect A restricted to SR – is empirically testable, provided that one uses the same measuring rods and clocks and conventions for simultaneity in the different inertial reference systems. In this sense, the principle of relativity of SR has been very well confirmed since 100 years. The principle of general covariance, however, which corresponds to aspect B, is not empirically testable; it can be satisfied by a mathematical technique (writing the equations in tensorial form) for the formulation of the laws.

General Relativity (GR)

(i) One important difference between aspects A and B can be expressed as follows. The idea of covariance in the sense of aspect B, i.e. coordinate independent formulation of the laws, is concerned with a *property of the formulation* of laws, whereas aspect A, i.e. invariance of laws under change of reference systems, is concerned with a *property of the laws* themselves.[100]

(ii) Neither an extended principle of relativity, in the sense that the laws of nature are invariant with respect to reference systems under arbitrary motion – which would correspond to aspect A for GR – nor a general covariance principle, in the sense of aspect B, are specific for GR. The first is not specific for GR because it is not generally satisfied by it, and the second is not specific for GR because general covariance of a theory can always be achieved. This can be further illustrated by the coordinate independent formulation of Newton's theory of gravitation by Cartan and others.

[100] This point was realised by Einstein (1924, ÜdÄ) p. 90: General covariance is "more characteristic of the mathematical form of this theory than its physical content". It was mainly Kretschmann's point. Pauli (1921, RTh), p. 187, stresses that the generally covariant formulation of the physical laws implies physical content only via the principle of equivalence and not by itself.

(iii) The general covariant formulation of physical laws such that they hold equally in all coordinate systems (according to B) does not imply a generalised relativity principle claiming a generalised equivalence of reference systems (according to A). This can be illustrated within the framework of classical, Lorentz-invariant mechanics, i.e. in a situation without gravitational fields. In an inertial system $I(x^\mu)$ with *Cartesian* coordinates x^μ, the equation of motion of a force-free mass point reads

$$\frac{d^2x^\mu}{d\tau^2} = 0 \, .$$

In an accelerated frame of reference, the equation of motion of the same mass point looks quite different, since various inertial forces are induced by the acceleration of the reference system. (Note that also in Newtonian mechanics the equation of motion of a force-free mass point $\ddot{x} = 0$ reads, e.g. in a rotating reference system with angular velocity $\vec{\Omega}$,

$$\vec{\ddot{x}}' = -2(\vec{\Omega} \times \vec{\dot{x}}') - (\vec{\Omega} \times (\vec{\Omega} \times \vec{x})) - (\vec{\dot{\Omega}} \times \vec{x})$$

and contains several inertial forces.) One can, however, obtain a unified formulation of the equation of motion, if the metric tensor and its derivatives are incorporated into the equation. The equation of motion, then, reads

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} = 0$$

independent of the coordinate system.[101] However, it is obvious that this covariant reformulation of the equation of motion does not change the observable inertial forces in any way. Quite the same situation can be found in general relativity, when the Minkowski spacetime is replaced by a Riemannian spacetime.

(iv) In the Riemannian spacetime of general relativity, there is an additional and more sophisticated argument which shows that reference systems and coordinate systems are not simply equivalent. On the level of cosmological solutions of Einstein's field equations, the large-scale structure of spacetime generally restricts the possibility of global coordinates. In general, an atlas of a global solution will consist of several partly disconnected local charts.[102] As a simple but interesting illustration, we consider a cosmological solution with incoherent matter. In this case, the

---

[101] Covariant formulations in Minkowskian spacetime of more general laws of physics can be found and are discussed in: Heintzmann, Mittelstaedt (1968, PGB).

[102] cf. Hawking, Ellis (1973, LSS) p. 11ff.

material elements which move freely on geodesics represent a velocity field $v_\mu(x^\lambda)$ that constitutes a cosmic frame of reference. We will imagine here, that at each material reference point there is an observer who is equipped with a standard clock, rods, and a radar system. Generally, a coordinate system attributes spacetime values to each reference point. If the system of coordinates is comoving, then the position coordinates are constant for each reference point. In analogy to inertial systems, it is of interest to find spacetime coordinates such that the clocks of all reference points are synchronised, e.g. by Einstein synchronisation with $\varepsilon = 1/2$. In this case, the totality of local clocks on reference points would constitute a cosmic time scale.

This is, however, not always possible. The covariant derivative of the cosmic velocity field v_μ is a tensor $v_{\mu;v}$ which can be decomposed in an irreducible way into three parts, the traceless symmetric shear tensor $\sigma_{\mu v}$, the tensor of expansion (dilatation), and the antisymmetric tensor $\omega_{\mu v}$ of rotation. Even if we assume – as usual in cosmology – that the shear tensor disappears, we find that a universal synchronisation – and thus, the existence of a cosmic time scale – is only possible if the rotation tensor $\omega_{\mu v}$ of the cosmic matter vanishes. If this is not the case,[103] we cannot define a universal cosmic time scale by means of which we can state that the age of the universe is almost 15 billion years.[104] Hence, we conclude that, within the context of Riemannian spacetime, for a given global frame of reference represented by a global velocity field v_μ there could be serious restrictions for systems of coordinates, i.e. for consistently attributing (x^k, t) values, to each reference point.

(v) Similar to what has been said above with respect to special relativity, the non-equivalence between aspects A and B is evident from the fact that general covariance, in the sense of coordinate independent formulation of laws (aspect B), is not empirically testable, because it can be achieved by a mathematical technique. Its actual role is rather a heuristic one.

On the other hand, the theory of general relativity, and this means the theory of gravitation expressed by Einstein's field equations, has successfully been confirmed by a great number of tests. It must be emphasised that these tests do not justify the covariance principle, which is a heuristical point of view, but Einstein's theory of gravitation. Not only the three

[103] The most famous example of a cosmological model with rotation is Gödel's solution of Einstein's equations.

[104] For more details cf. Mittelstaedt (1989, ZBP) p. 159–160; and Misner, Thorne, Wheeler (1973, Grav) p. 715, who call coordinates with a cosmic time scale "synchronous".

famous original predictions about the perihelion of Mercury, the deviation of light rays, and the gravitational red shift (tests have been repeated with most recent and accurate methods), but also numerous further ones which include the equivalence of gravitational and inertial mass, clock comparisons, the Thirring–Lense effect, gravitational lenses, black holes, and gravitational waves.[105]

6.5 Reply to the Objections

6.5.1 (to 6.1.1) According to the proposed answer (Sects. 6.4.3–6.4.7), different levels of invariance have to be distinguished. The first three kinds of invariance – under internal transformations, under inertial movement I (Galilean transformations), and under inertial movements II (Lorentz transformations) – abstract from acceleration and from any gravitational fields. For accelerated frames of reference and for reference frames in the presence of gravitational fields, there are no further invariances of the physical laws. 6.5.2 (to 6.1.2)

According to the objection, there is a difference between coordinate systems as mathematical structures and reference frames like inertial systems or accelerated systems. Hence, invariance under changes of coordinate systems (as mathematical structures) does not in general imply invariance under changes of reference frames. By contrast – since according to 6.3.1 every reference frame is also a physical system, conjoint with a local coordinate system – invariance with respect to a reference system implies invariance with respect to the conjoint local coordinate system. 6.5.3 (to 6.1.3) It states correctly in

this objection that not all laws of nature are invariant with respect to mirror reflection, i.e. “nature” can distinguish between left and right. But, as we explained in Sect. 6.4.2, spacetime invariance is concerned with real continuous displacement and movement of (physical) reference frames (including local coordinate structures) and not with virtual movements or conceptual changes alone. Now, since no real (physical) reference system can turn by mirror reflection, invariance or symmetry with respect to mirror reflection has been listed under “Discrete Symmetries” (recall 5.3.3). Spacetime invariance, however, is concerned with continuous translation or rotation in space, or displacement of time, or inertial or arbitrary movement of the reference frame. Therefore, the conclusion of this objection is not proved and nothing hinders that every law of nature is spacetime invariant in this sense.

6.5.4 (to 6.1.4) The first premise of this objection is false: Spacetime invariance does not imply that the displacement on the time axis (coordinate) is the same as the displacement along one of the spatial coordinates. As long as the laws are invariant under some displacement on the time coordinate, together

[105] For a discussion of experimental tests for GR see Shapiro (1980, ECP) and Sexl, Urbantke (2002, GRK) Chap. 4.

with some displacement on the spatial coordinates (at least one of them), we may speak of spacetime invariance. Thus, it is perfectly compatible with spacetime invariance that a reference system can move only in the direction of increasing time, though in the directions of increasing or decreasing (of values) on the spatial coordinates (as Wigner points out in the quotation). Therefore, the conclusion in objection 6.1.4 is not proved.[106] 6.5.5 (to 6.1.5 and 6.1.6)

The main premise used in both objections 6.1.5 and 6.1.6 is that spacetime invariance implies invariance with respect to time reversal. This premise is false, since spacetime invariance means invariance with respect to continuous translations of spacetime, and this invariance does not imply invariance with respect to time reversal. According to 6.4, there are three kinds of invariance to be distinguished. But spacetime invariance implies in none of them invariance with respect to time reversal.

(i) In the first way, spacetime invariance means that the respective laws are invariant under internal transformation of reference systems. But, as was explained in 6.4.3, an internal transformation (here, with the focus on time) only means replacing some value of the time coordinate by another one. Or – in Weinberg's words (cf. 6.4.3) – to set your clocks differently in your laboratory. But this does not involve any direction on the time axis; it involves only continuous order. Thus, although such a continuous order is compatible with time reversal, it does not imply it.

This can be illustrated by a simple example: In Newtonian mechanics, the equation of motion with friction

$$m\frac{d^2x}{dt^2} = -\alpha\frac{dx}{dt}$$

is invariant with respect to the continuous translations of spacetime

$$x \to x' = x + a\ , \quad t \to t' = t + b\ ,$$

but this equation is not invariant with respect to time reversal $t \to t' = -t$.

(ii) In the second and third ways, spacetime invariance means that the respective laws are invariant under transformations of inertial systems, either with Galilean transformation or with Lorentz transformation (cf. 6.4.4 and 6.4.5). But this second way of spacetime invariance cannot supply invariance with respect to time reversal. This can be seen as follows: As it was explained in 6.4.2, spacetime invariance is concerned with *real* continuous movements of the reference frame. But mirror reflections or time reversal are *virtual* movements. This is connected with the following point of understanding the movement of a coordinate system (or of a particle of a coordinate system) which was stressed by Wigner: The moving system

[106] For the question of invariance with respect to time reversal see the commentary to the next objections 6.1.5, 6.1.6 and Sect. 7.2.3.4.3.

can cross a value (line) on a spatial axis in both directions, but on the time axis only in the direction of increasing t (recall objection 6.1.4 and the reply to it in 6.5.4). Such a movement in the direction of increasing t can be without a connection to an entropy increasing process (for example in planetary movement), or it can be connected to an entropy increasing process (as is the case in thermodynamic or cosmological processes). In the first case, there is compatibility with time reversal of the laws, but, of course, no implication of time reversal, which only concerns virtual movement or change of the (mathematical) coordinate system. In the second case, the question of the compatibility is a difficult problem and depends very much on how strong the so-called "irreversibility" of entropy increasing processes is interpreted.[107]

This can be illustrated by the following example: The Newtonian equation of motion $m\frac{d^2x}{dt^2} = \alpha t$ is invariant under the Galileo transformation $x \to x' = x - vt,\ t \to t' = t$, but this equation is not invariant with respect to time reversal $t \to t' = -t$.

6.5.6 (to 6.1.7) The first premise of objection 6.1.7 is based on a confusion: It is not the law that determines the geometry of spacetime, but the distribution of matter; and this fact is described by the law, i.e. by the field equations. Nothing hinders, therefore, that the respective laws, which describe this fact are spacetime invariant. And hence, the conclusion in objection 6.1.7 is not proved. To be a little more detailed:

Einstein's field equations

$$R^{\mu v} - \frac{1}{2}Rg^{\mu v} = \kappa T^{\mu v} \quad (\kappa \text{ is the relativistic gravitational constant})$$

connect the metric $g_{\mu v}$ of spacetime and its first and second derivatives with the distribution of matter, i.e. the energy momentum (matter) tensor $T^{\mu v}$. These equations (10 coupled non-linear partial differential equations for the 10 coefficients of the metric tensor) describe the dependence of the structure of spacetime (represented by the metric tensor) on the distribution of matter (represented by the matter tensor). However, the covariant derivative of each the left hand side of Einstein's field equation vanishes identically, i.e.

$$\left(R^{\mu v} - \frac{1}{2}Rg^{\mu v}\right); v = \kappa T^{\mu v}; v = 0\ .$$

Hence, the field equations provide only six independent differential equations for the metric.

This is just the correct number of equations to determine the metric $g_{\mu v}$, since four of its ten components can have arbitrary values, according to the

[107] On the rather weak interpretation – as a very small probability of recurrence – which will be proposed in Chap. 7, there is hardly a serious incompatibility.

four degrees of freedom that come from coordinate transformations. In this way, the distribution of matter and the boundary conditions produce what Weyl called the "guiding field" that makes the geometrical structure of spacetime curved, i.e. transforms the trajectories of the Minkowskian spacetime into geodesics of a Riemannian spacetime.

The sometimes used slogan mentioned in this connection:

> "Space acts on matter, telling it how to move.
> In turn matter reacts back on space, telling it how to curve",[108]

is only half true:

First, there is no such symmetry[109] as the slogan suggests; if there is matter, then there is a gravitational field; but the converse is not claimed by GR. If there is a gravitational field, there need not to be matter. Indeed, in case of $R_{\mu v} = 0$ there are Riemannian spacetime structures without field producing matter.

To further elucidate this point, it is convenient to decompose the Riemannian curvature tensor $R^{\alpha}_{\beta\gamma\delta}$ according to[110]

$$\begin{aligned} \mathrm{C}_{\alpha\beta\gamma\delta} = R_{\alpha\beta\gamma\delta} &+ \frac{1}{n-2}(g_{\alpha\gamma}R_{\beta\delta} - g_{\alpha\delta}R_{\beta\gamma} + g_{\beta\delta}R_{\alpha\gamma}) \\ &+ \frac{R}{(n-1)(n-2)}(g_{\alpha\delta}g_{\beta\gamma} - g_{\alpha\gamma}g_{\beta\delta}) \end{aligned}$$

where $\mathrm{C}_{\alpha\beta\gamma\delta}$ is the Weyl tensor, which describes the curvature of the matter free spacetime.

Second, the term "geometry" in the slogan is misleading because the curved spacetime of GR is not a purely mathematical–geometrical coordinate system but a physically determined Riemannian spacetime metric (recall 6.3).

Third, the field equations obey two fundamental correspondence conditions, which show that the slogan is not correct in the special relativistic *and* in the non-relativistic limit:

(i) If gravitation decreases towards 0, then the field equations contain SR as a limiting case. This is possible in two ways, globally and locally: In the limit of the vanishing gravitational field (curvature) a reduction to a global inertial frame can be introduced everywhere in which SR holds generally. In the limit of the vanishing gravitational field (curvature) locally, a local inertial frame is produced in which the laws of physics take their special

---

108 cf.: Misner, Thorne, Wheeler (1973, Grav) p. 5 and 408.or: Resnick and Halliday (1985, BCR) p. 296. We do, however, not claim that these authors wanted to be interpreted rather literally, concerning those passages including the slogan: In Misner, Thorne and Wheeler they occur in introductory chapters. However, we take this opportunity to clarify an important point.

109 Recall Sect. 6.4.7.2(5).

110 cf. Hawking et al. (1973, LSS), p. 41.

relativistic form. There are, however, in both reductions no restrictions on the metric, although there are severe constraints on matter and field.

(ii) In case of very weak gravitation, very small pressures and very low velocities, GR reduces to Newton's theory of gravitation as a limiting case.[111]

[111] cf. the elaboration of the Newtonian limit in Ehlers (1991, NLG).

7

Dynamical and Statistical Laws

7.1 Are all Laws of Nature Either Dynamical or Statistical Laws?

7.1.1 Arguments Contra

7.1.1.1 If all laws of nature are either dynamical or statistical laws, then all phenomena systematically observed and described in physics can be described and explained by one or the other type of laws. Now, since about twenty years chaotic phenomena belonging to the so-called dynamical chaos have been systematically observed and described. But these phenomena cannot be explained or described by either dynamical or statistical laws. Therefore – under the additional metascientific assumption that these phenomena are not law-less – it does not hold that all laws of nature are either dynamical or statistical laws.

7.1.1.2 If all laws of nature were either dynamical or statistical laws, then there would be no third type of law which could describe the same phenomena in one area in an alternative but also sufficiently complete way like these two types of laws. But the principles of "least action",[1] which amount to minimising a certain functional called "action"

$$\mathbf{S} = \int_{t_1}^{t_2} L(x_k, x'_k, t)dt$$

with a Lagrangian L, are of a third type of law that can describe the respective phenomenon in a completely alternative way. Therefore, not all laws of nature are either dynamical or statistical laws.

[1] For the approach with principles of least action see Feynman (1964, LPh), Vol. II, Chap. 19. For historical investigations on the principle of least action see Stöltzner, Weingartner (2005, FTK).

7.1.2 Arguments Pro

Science is only possible if there are laws of nature. But, as Aristotle says, science is possible where we have laws or regularities which say that something happens always (without exceptions) or in most cases (with a few exceptions): "But that there is no science of the accidental is obvious; for all science is either of that which is always or of that which is for the most part."[2] Now, Aristotle's description "to happen always (without exceptions)" and "to happen in most cases (with a few exceptions)" can be interpreted in modern terms by saying: it happens according to dynamical laws, or it happens according to statistical laws. Therefore – if Aristotle is right – the existence of (natural) science, which is a fact, implies that the laws of nature (used in science) are either dynamical or statistical laws.

7.1.3 Proposed Answer

All laws of nature known so far are either dynamical or statistical laws. This can be shown by an exhaustive list of types or groups of laws of nature known so far. Since in all the different types or groups we find either dynamical or statistical laws:

(1) The laws of CM (classical mechanics). These laws have the form of differential equations and belong to the group of dynamical laws.
(2) The laws of CEM (classical electrodynamics; also called Maxwell's equations). These laws have also the form of differential equations and belong to the group of dynamical laws.
(3) The laws of SR (special theory of relativity). These laws are the laws of CEM and the corrected versions of CM (corrected with the help of the Lorentz transformations). Therefore, they also belong to the group of dynamical laws.
(4) The laws of GR (general theory of relativity). Einstein's general field equations (and further developments of it) also belong to the group of dynamical laws.
(5) The laws of thermodynamics, like the law of entropy, and the laws of radiation, like Planck's law of radiation, belong to the group of statistical laws.
(6) The laws of QM (quantum mechanics) are of both types. Those like the Schrödinger equation are dynamical laws; those which make predictions about the outcomes of measurement processes are statistical laws.
(7) The (most universal) laws of chemistry, like the general law of gases and van der Waals' equation of real gases, are statistical laws.
(8) The (most universal) laws of biology, like the law of Hardy–Weinberg, which is a further development of Mendel's laws, are statistical laws.[3]

[2] Aristotle (Met), 1027a20. cf. (Phys) 198b34.

[3] This list is, of course, restricted to the laws of natural sciences. This restriction is one which concerns this book, which, as was pointed out in the preface, deals with

Since this list is more or less exhaustive with respect to laws of nature known so far, all laws of nature known so far are either dynamical or statistical laws.

7.1.4 Reply to the Objections

7.1.4.1 (to 7.1.1.1) Though chaotic motion (belonging to dynamical chaos)[4] is not law-less, its trajectories cannot be described or explained or predicted with sufficient accuracy. First, chaotic motion is not law-less because it is generated by physical systems obeying dynamical laws, in the sense that these underlying laws determine uniquely the time evolution of a certain state of the system from its previous states. On the other hand, its trajectories cannot be described or explained or predicted with sufficient accuracy (in many cases with no accuracy at all) because the system is non-linear such that smallest and closest points (or trajectories) in the initial state (initial conditions) separate exponentially fast in a bounded region of phase space.

Therefore, the first premise in the objection is not entirely correct, since not all phenomena – here, especially those of non-linear systems – can be described and explained to a sufficient degree of accuracy by one or the other type of law. Consequently, the conclusion of the objection is not proved by this argument.

7.1.4.2 (to 7.1.1.2) The principle of *least action.* In order to elucidate this problem in more detail within the framework of classical mechanics, we consider the trajectory $x_k(t)$ of a given body under the influence of forces or fields in the time interval $t_1 \leq t \leq t_2$. This trajectory is a real and measurable phenomenon which allows, however, for two alternative interpretations.

(a) The *causal* interpretation
Starting from the initial values $x(t_1)$, $x'(t_1)$, the trajectory is created step by step, ending up with the final values $x_k(t_2)$ and $x'_k(t_2)$. A law which describes this subsequent continuous creation under the influence of forces must hold irrespective of the initial conditions. A law of this kind is given by Newton's classical equation of motion

$$\frac{d^2 x_k}{dt^2} = F_k(x_l, x'_l, t)$$

since this differential equation implies – together with the initial values mentioned – the trajectory $x_k(t)$.

(b) The *teleological* interpretation
The trajectory $x_k(t)$ minimises (or maximises) the "action"

laws of nature described and discovered by the natural sciences, and especially by physics.

[4] There are other types of chaotic motion, like those belonging to quantum chaos. For more on both types of chaos see Sect. 9.4.

$$\mathbf{S} = \int_{t_1}^{t_2} L(x_k, x'_k, t)dt \ ,$$

where L is a convenient Lagrangian of the problem considered. Here, the initial and final values are kept constant, and the trajectory is varied and chosen such that the action S assumes an extreme value. The law which expresses this *global* requirement is the variational principle

$$\delta \int_{t_1}^{t_2} L(x_k, x'_k, t)dt = 0$$

with fixed boundaries t_1 and t_2. Elaborating this variation, one arrives again at the trajectory $x_k(t)$.

Note that the trajectory $x_k(t)$ is a real phenomenon which can be measured and which belongs to the external reality. Neither the splitting of the trajectory into initial conditions and a differential equation[5]

$$\frac{d^2x_k}{dt^2} = F_k(x_l, x'_l, t)$$

nor the variational principle $\delta S = 0$ are observable in the same sense as the trajectory. They are motivated by methodological arguments and laden with non-empirical interpretations. Newton's equation of motion expresses the causal interpretation of the trajectory, whereas the variational principle is the formal expression of the teleological interpretation. It is easy to see that the variational principle $\delta S = 0$ implies the differential equation

$$\frac{d^2x_k}{dt^2} = F_k(x_l, x'_l, t) \ .$$

The converse is, however, not the case, since neither the Lagrangian nor the action is completely determined by the trajectory. There are, in general, several equivalent formulations of an action principle. This plurality of empirically equivalent action principles shows that the well-known metaphysical interpretation of the action principle is untenable. A mechanical process does not proceed *such that* at its end the used amount of time, of energy, or of action is minimal.[6] Physical nature has no goals in the sense as living organisms have goals. Instead, the large variety of action principles for the same process clearly demonstrates that many different quantities can be minimised or maximised by a given mechanical process. Hence, there is no definite tendency or strategy.

---

[5] For this splitting cf. 11.2.3 and 11.2.2.1.

[6] The idea that an action principle is the mathematical expression of some optimisation program was first conceived by Leibniz, and illustrated by Snell's law for the refraction of light and by Fermat's principle. cf. Leibniz (Met), Sect. 22.

In many fields of physics, both interpretations can be applied. This means that, for a given "equation of motion" which represents a causal interpretation, a variational principle can be found which represents a teleological interpretation and implies the equation of motion. In classical mechanics, relativistic mechanics, and electrodynamics combined with relativistic mechanics both methods are well known.[7] Even in quantum mechanics both methods can be applied. Originally, quantum mechanics was formulated by Schrödinger's time dependent differential equation in accordance with a causal interpretation of the wave function $\psi(x_k, t)$. For a long time, an action principle could not be found, since in its standard version, quantum mechanics does not use the Lagrangean formalism. In 1942 Feynman succeeded in reformulating quantum mechanics in terms of a Lagrangean formalism and an action principle.[8] In this formalism, all the trajectories – called "paths" here – are completely fictitious and not observable, since trajectories in the classical sense do not exist in quantum mechanics. They serve merely as a means for formulating the time-dependent wave function $\psi(x_k, t)$.[9] The difficulties to reformulate quantum mechanics in terms of an action principle indicate that it is perhaps not possible to find a corresponding action principle for every law that is expressed by a differential equation. However, for several classes of differential equations it can be shown that for a given differential equation there are action principles which imply the differential equation in question. In particular, for any ordinary, second order differential equation there are *infinitely many* corresponding action principles.[10]

7.2 Is One Type of Law Reducible to the Other?

7.2.1 Arguments Pro

7.2.1.1 If all laws of nature are either dynamical or statistical laws, then neither of these classes is empty. But, if all statistical laws can be ultimately reduced to dynamical laws, then the second class (of statistical laws) is empty. Now, there seem to be good reasons for reducing all statistical laws to dynamical laws. Thus, Planck says:

> "I believe and hope that a strict mechanical significance can be found for the second law along this path, but the problem is obviously extremely difficult and requires time."[11]

[7] For a survey about these methods cf. F. Rohrlich (1965, CCP).

[8] R.P. Feynman, PhD thesis, Princeton 1942, published in: Rev. Mod. Phys. 20, p. 367 (1948).

[9] For all details about the path-integral method and its application in many fields of physics, we refer to the monograph by Kleinert (1993, PDI), and the literature quoted there.

[10] Bolza (1909, VVR), pp. 37–38. Courant, Hilbert, (1968, MMP), p. 219.

[11] Planck in a letter to his friend Leo Graetz. Cited in Kuhn (1978, BBT), p. 27.

> "The principle of energy conservation requires that all natural occurrences be analysable ultimately into so-called conservative effects like, for example, those which take place in the motion of a system of mutually attracting or repelling material points, or also in completely elastic media, or with electromagnetic waves in insulators. ... On the other hand, the principle of the increase of entropy teaches that all changes in nature proceed in one direction. ... From this opposition arises the fundamental task of theoretical physics, the reduction of unidirectional change to conservative effects."[12]

Therefore, there seem to be good reasons that all laws of nature are dynamical laws.

7.2.1.2 Statistical laws do not describe (determine) the particular (individual) details of the state of the physical system. Thus, statistical laws express a lack of knowledge of the details of a state (of the individual physical system). But, as Laplace says, an intelligence who knows all the details of one state of a (physical) system (or of the whole universe), together with all laws of nature, could predict and retrodict all the other states of the system (of the universe):

> "We ought to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it – an intelligence sufficiently vast to submit these data to analysis – it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes."[13]

Thus, in principle – that is, if all the details of a certain state of the system are known – one could predict and retrodict all other states, as Laplace says, and, consequently, all statistical laws would turn into dynamical laws. Therefore – in principle, i.e. if there is no lack of knowledge concerning the details of a state (or if this lack can be sufficiently diminished)[14] – all statistical laws are reducible to dynamical laws.

[12] Planck in a paper read to the Prussian Academy of Science in 1897. Cited in Kuhn (1978, BBT), p. 28.

[13] Laplace (1814, EPr), Chap. 2.

[14] It has to be noted that it is important in this connection that Laplace's intelligence (who knows the state in all details) must not belong as a part to the physical system (having these states). If the physical system is the whole universe, this intelligence has to be immaterial and has to be outside the universe. This has been shown by Breuer (1995, IAS) and Breuer (1996, SDQ).

7.2.2 Argument Contra

All dynamical laws are time reversible; i.e. in a differential equation like that of Newton's second law of motion (for CM) or that of Schrödinger for QM, one can replace the sign t (for time) by $-t$ without making the law invalid.[15] But this is not possible in statistical laws, like those of thermodynamics or those describing processes of radiation.

Therefore, neither can be reduced to the other.

7.2.3 Proposed Answer

Neither statistical laws can be reduced to dynamical laws nor dynamical laws can be reduced to statistical laws. The reason is that both types of laws have several properties which differ in such a way that a reduction is impossible. This will be shown in the following three parts: First, some historical remarks concerning the two types of laws will be given (7.2.3.1). Secondly, it will be shown in detail, by comparing the properties of dynamical and statistical laws, that neither type of law can be reduced to the other (7.2.3.2, 7.2.3.3, 7.2.3.4). Thirdly, it will be shown that, though a reduction is impossible, both types of laws are nevertheless compatible in such a way that they may hold in the same physical system (7.2.3.5).

7.2.3.1 Historical Introduction

The world view underlying Laplace's quotation (7.2.1.2 above) was based on the belief that all physical systems are – if analysed in their inmost structure – ultimately mechanical systems. Since a clock was understood as a paradigm example of a mechanical system, the main thesis of the mechanistic world view could be expressed by saying that all complex systems (things) of the world – even most complicated ones like gases, swarms of mosquitoes, or clouds – are ultimately (i.e. if we would have enough knowledge of the detailed interaction of the particles) clocks. Or, put in words Popper used in his A.H. Compton Memorial Lecture: "All clouds are clocks".[16]

After the discovery of statistical laws in thermodynamics and later in other areas, there was a general doubt with respect to the mechanistic and deterministic interpretation of the world. At this point, a clarification of the concept "thermodynamics" seems to be suitable. In general, thermodynamics is a discipline which is concerned with the question how the properties of material (physical) systems change with temperature. This subject matter can be investigated in a twofold way. First, it can be studied only on the macroscopic scale. In this case, one is concerned solely with relations between macroscopic observable quantities and does not look at (or, historically: was

[15] For QM, this was shown by Wigner (1932, OZQ) and Dirac (1937, ROQ).
[16] Popper (1965, CaC), p. 210.

ignorant of) an underlying deeper (micro)structure. This approach, which may be called classical thermodynamics, is "phenomenological" (recall Chap. 2) in the sense that its answers are, to some extent, independent of an underlying and explaining structure. However, they still provide boundary conditions that have to be satisfied by any microscopic model about the underlying explanatory structure. Second, this discipline can be studied at the microlevel, with the focus to use the behaviour at the microlevel as the explanatory structure of the behaviour at the macrolevel. This approach was developed in two forms: As the kinetic theory of gases by Clausius and Maxwell, and as the statistical mechanics discovered and invented by Boltzmann. Whereas the first approach can be (and is) done rather independent of the second, the second is always understood as intrinsically connected with the first, since the behaviour on the microlevel is interpreted as the underlying structure explaining the behaviour on the macrolevel. If not indicated otherwise, "Thermodynamics" will subsequently be understood in the second sense.[17]

That there are physical truths which are statistical in character was clear for Boltzmann and for Poincaré, who both underline the importance of Maxwell's, Clausius', Gibb's and Carnot's discoveries:

> "Schon Clausius, Maxwell u.a. haben wiederholt darauf hingewiesen, daß die Lehrsätze der Gastheorie den Charakter statistischer Wahrheiten haben. Ich habe besonders oft und so deutlich als mir möglich war betont, daß das Maxwellsche Gesetz der Geschwindigkeitsverteilung unter Gasmolekülen keineswegs wie ein Lehrsatz der gewöhnlichen Mechanik aus den Bewegungsgleichungen allein bewiesen werden kann, daß man vielmehr nur beweisen kann, daß dasselbe weitaus die größte Wahrscheinlichkeit hat und bei einer großen Anzahl von Molekülen alle übrigen Zustände damit verglichen so unwahrscheinlich sind, daß sie praktisch nicht in Betracht kommen."[18]

Poincaré, after commenting on Carnot's principle and discussing irreversible processes which cannot be explained with the help of classical mechanics, gives an example:

> "A drop of wine falls into a glass of water; whatever may be the law of the internal motion of the liquid, we shall soon see it colored of a uniform rosy tint, and however much from this moment one may shake it afterwards, the wine and the water do not seem capable of again separating. Here, we have the type of the irreversible physical phenomenon: to hide a grain of barley in a heap of wheat, this is easy; afterwards to find it again and get it out, this is practically impossible. All this Maxwell and Boltzmann have explained; but the one who has seen it most clearly, in a book too little read because it is a little

[17] cf. the lucid exposition of thermodynamics in Longair (1984, TCP), Chaps. 6 and 7.

[18] Boltzmann (1896, EWB), p. 567.

difficult to read, is Gibbs, in his 'Elementary Principles of Statistical Mechanics'."[19]

One of the first philosophers who noticed that a certain imperfection in all "clocks" allows to enter chance and randomness was Charles Sanders Peirce:

> "But it may be asked whether if there were an element of real chance in the universe it must not occasionally be productive of signal effects such as could not pass unobserved. In answer to this question, without stopping to point out that there is an abundance of great events which one might to be tempted to suppose were of that nature, it will be simplest to remark that physicists hold that the particles of gases are moving about irregularly, substantially as if by real chance, and that by the principles of probabilities there must occasionally happen to be concentrations of heat in the gases contrary to the second law of thermodynamics, and these concentrations occurring in explosive mixtures, must sometimes have tremendous effects."[20]

The question was now: Could it not be the case that all laws are statistical and the deterministic outlook is only on the surface of macroscopic phenomena? That is, all complex systems (things) of the world are in fact – in their inmost structure, i.e. on the atomic level – like gases or swarms of mosquitoes or clouds. This led to another extreme picture discussed by Popper: "All clocks are clouds".[21]

The question whether all physical laws can be reduced to or based on statistical laws was, however, not a serious topic at the time of Poincaré and Boltzmann. One reason for that was that quantum theory was not yet available. Rather there were two important questions:

(1) Are physical laws which are statistical, like the second law of thermodynamics (the law of entropy), compatible with the basic dynamical laws (of classical mechanics)?
(2) Are the statistical laws, like the law of entropy, explainable with the help of (or reducible to) dynamical laws?

Zermelo thought to have proved that the answer to (1) is negative. But Boltzmann explains the misunderstandings of Zermelo, and shows that there is no incompatibility.[22] Planck hoped that (2) is true and stresses that he does not go as far as Zermelo, who was Planck's assistant at this time:

---

[19] Poincaré (1958, VSc), p. 97. cf. the section on Poincaré's recurrence theorem 7.2.3.4.3(2c).
[20] Peirce (1960, CPC), Chap. 6.47.
[21] Popper (1965, CaC), ibid.
[22] Boltzmann (1896, EWB) and (1897, ZAM). For more details on that see Sect. 7.2.3.4.3(2) below.

> "Zermelo, however, goes farther [than I], and I think that incorrect. He believes that the second law, considered as a law of nature, is incompatible with any mechanical view of nature. The problem becomes essentially different, however, if one considers continuous matter instead of discrete mass-points like the molecules of gas theory. I believe and hope that a strict mechanical significance can be found for the second law along this path, but the problem is obviously extremely difficult and requires time."[23]

But neither of these extreme pictures – reduction to dynamical laws "all clouds are clocks" or reduction to statistical laws "all clocks are clouds" – proved satisfactory as an explanation of everything. The heroic ideal to explain everything by one (or one kind of) principle[24] had to be replaced by the aim to find relatively few (kinds of) principles (laws) for relatively many facts. This more modest ideal was manifest already in Greek Science, especially in Aristotle's theory of science in his posterior analytics, and in Euclid's Elements. Feynman expresses it for the situation today in the following words: "We do not have one structure from which all is deduced, we have several pieces that do not quite fit exactly yet."[25] At the turn of the century and in the first half of it, many physicists accepted a view which can be roughly stated as follows:

With respect to some areas (mainly macroscopic), deterministic laws with good predictability for single events give an adequate description and explanation.

With respect to other areas (thermodynamics, friction, diffusion, radiation and microscopic areas), statistical laws with no good predictability for the single event but with predictability for the whole aggregate give an adequate description and explanation.

7.2.3.2 Properties of Dynamical Laws

The dynamical law describes the time development of a physical system S in such a way that the following condition D1 is satisfied:

- D1 The state of the physical system S at any given time t_i is a definite function of its state at an earlier time t_{i-1}. A unique earlier state (corresponding to a unique solution of the differential equation) leads under the time evolution to a unique final state (again corresponding to a unique solution of the equation).
- D2 Condition D1 is also satisfied for every part of the physical system, especially for every individual body (object) as part of the system, even if the

[23] Planck in a letter to his friend Leo Graetz. Cited in Kuhn (1978, BBT), p. 27.

[24] This was an ideal of Descartes' philosophy: all factual truths should be derived from one axiom, the "Cogito ergo sum".

[25] Feynman (1967, CPL), p. 30.

individual objects may differ in the classical or in the quantum mechanical sense.[26]

Dynamical laws have been applied successfully (i.e. such that the laws were confirmed by the application) to those physical systems which satisfy the following further conditions D3 and D4:

D3 The physical system S is periodic, that is, the state of S repeats itself after a finite period of time and continues to do so in the absence of external disturbing forces.

D4 The physical system S has a certain type of stability which obeys the following condition: Very small changes in the initial states, say within a neighbourhood distance of ε, lead to proportionally small (no more than in accordance of a linearly increasing function of time) changes $h(\varepsilon)$ in the final state. This kind of stability, which survives small perturbations and leads to relaxation afterwards, is called *perturbative stability* and holds in many linear systems.[27]

D1 is the main condition for dynamical laws. It is represented in the quotation of Laplace (7.2.1.2). This quotation, moreover, expresses Laplace's view that the dynamical laws are of global generality and applicability; they are *the* laws governing the whole universe. This global idea of Laplace can be illustrated by the following picture:

Assume a film is made of the world, i.e. of the events happening in the whole universe. After the film is developed, we cut it into pieces corresponding to single film-pictures. Now, we put the single pictures successively in time (in the order of time) into a long card index box, like the cards of a library catalogue. Then, one special state of the universe at a certain time t corresponds to one such card (film picture) of the catalogue. One can follow one trajectory across the (perpendicular to the) catalogue cards.

Interpreted with the help of this illustration, Laplace's idea expressed in the quotation means that it suffices to know the law(s) of nature and one single catalogue card (film picture) corresponding to one state (of the universe) at a certain time t in order to construct all other cards of the catalogue, i.e. to predict and to retrodict all the other states of the universe. D1 is usually taken as the defining condition for determinism.[28] This is correct only if not further properties, like predictability or conservation of information, etc. are thought to be implied by this kind of determinism, because D1 alone does not guarantee predictability or conservation of information, especially if D4 is not satisfied (see below).

Concerning condition D2, we have to observe that "physical system" in D1 can be understood in two ways: First, in a sense in which it was originally

[26] For objects in the classical or in the quantum mechanical sense see Sects. 10.2 and 10.3.

[27] cf. the discussion of the conditions D1, D3 and D4 in Holt, Holt (1993, RND).

[28] For more on that, in connection with causality, see Chap. 9.1.

understood in the history of physics, from the time of Newton on: As a macroscopic system, like the planetary system, which is composed of subsystems (say the planets or parts of the planets) in a definite sense, such that D1 holds also for all these subsystems, including individual bodies (objects).[29] In this sense, the dynamical law describes also the time development of the individual body or particle or object as a part of the time development of the whole physical system. This presupposes a concept of "individual object" that is unique and re-identifiable through time, as it is understood in CM.[30] This is, in fact, what is said by condition D2.

Secondly, "physical system" in D1 can be understood in such a way that the system does not consist of a definite composition of subsystems that are individual objects in the classical sense of object above. If dynamical laws are applied in QM, the (quantum) objects as parts (now not composable by Boolean operations only) of the physical system are not complete, and have to have commensurable properties.[31] D3 is not a necessary condition for the application of dynamical laws obeying D1 and D2, though D3 is satisfied in most cases where dynamical laws are applied. The main point is that, according to D3, there is recurrence of the state of the physical system after some finite period of time.

Are there important cases of physical systems which satisfy D1 but neither D3 nor D4? The answer to this question is: Yes. The systems in question are systems which show chaotic behaviour (or systems in chaotic motion).

Chaotic behaviour is non-periodic. And this holds also without any external disturbance. A consequence of that is a further characteristic of chaotic motion: The Poincaré map shows space-filling points. This is a method introduced by Poincaré which considers the points in which the trajectory cuts a certain plane. If the motion is chaotic, there will be no immediate recurrence, that is, the plane will always be cut at new points and, as time goes on, will be filled with points. But if the phase space is small, there will be recurrence of the trajectory after some finite period of time. To give an illustration: skiing in fresh powder snow is a great pleasure. But if the slope is small and one is skiing down frequently, the slope will be filled with traces, and after some time no new space is left, and thus, one has to use one's own traces again (recurrence). If the system is Hamiltonian and area preserving (finite region), then the Poincaré recurrence theorem holds. It says that the trajectory returns to a given neighbourhood of a point an infinite number of times (if the time is infinite or sufficiently long). If it is ergodic, then the

[29] According to the *Principia* of Newton for all extended material bodies.

[30] For a detailed discussion of these objects of classical physics see Chap. 10, especially Sects. 10.2.1.3 and 10.3.2. For the deeper reason of the classically motivated division into subsystems see also Zurek (1994, PSS), Sect. 11.2.

[31] See the detailed discussion in Chap. 10, especially Sect. 10.3.2.

system explores the entire region of phase space and eventually covers it uniformly (this implies also recurrency). In stronger kinds of chaotic motion, the trajectory might not cover the whole phase space and neither stay in a local area. The description is, then, more complicated.

For other important parameters, recurrency does not hold in chaotic motion. For example, phase density is non-recurrent, it will never come back to its initial state, independently of the direction of time. Thus, we have non-recurrency and time reversibility (the latter also for the relaxation property). In consequence, it is important to notice that non-recurrency and time irreversibility are not equivalent notions.[32]

The non-periodicity can also be measured by the invariant density, which measures how the iterations become distributed over the unit interval, and by the correlation function $f(m)$, which measures the correlation between iterations which are m steps apart.[33]

D4 was a hidden assumption of CM until the end of the 20th century. In other words, the laws of CM were understood in such a way that D4 is always satisfied. The neglect is expressed by Lighthill as follows:

> "Here I have to pause, and to speak once again on behalf of the broad global fraternity of practitioners of mechanics. We are all deeply conscious today that the enthusiasm of our forebears for the marvellous achievements of Newtonian mechanics led them to make generalisations in this area of predictability which, indeed, we may have generally tended to believe before 1960, but which we now recognise were false. We collectively wish to apologise for having misled the general educated public by spreading ideas about the determinism of systems satisfying Newton's laws of motion that, after 1960, were to be proved incorrect."[34]

On the other hand, that there are cases which violate D4, that is, where small initial deviations lead to unproportional (exponentially increasing) effects, was known from antiquity. A first warning, with respect to a principle like D4 in the area of epistemology or methodology, we find already in Aristotle: "the least initial deviation from the truth is multiplied later a thousandfold."[35] Kepler was convinced that the proportions of the distances between the planets and the sun (and their mutual distances) contribute to the harmony (in our interpretation: stability) of the whole planetary system. A specific warning with a counterexample is due to Maxwell:

---

[32] See Chirikov (1996, NLH), Sect. 2.2, and below Sect. 7.2.3.4.2(2).

[33] See Sect. 9.4, where more properties of chaotic motion are discussed.

[34] Lighthill (1986, RRF), p. 38.

[35] Aristotle (Heav), 271b8. The exponentially increasing error can be explained by the so-called Henon attractor. cf. Sect. 11.1.3.5(3) and Weingartner (1996, UWT), p. 58.

> "There is another maxime which must not be confounded with that quoted at the beginning of this article,[36] which asserts 'That like causes produce like effects'. This is only true when small variations in the initial circumstances produce only small variations in the final state of the system. In a great many physical phenomena this condition is satisfied; but there are other cases in which a small initial variation may produce a very great change in the final state of the system, as when the displacement of the 'points' causes a railway train to run into another instead of keeping its proper course."[37]

Experienced highlanders in mountainous countries like Tyrol know very well that extremely small events can lead to a bursting of an avalanche which might destroy huge forests and even a city.

It should be noted that the unproportional effects need not to be chaotic. In the example of Maxwell, the running of the train in a different direction is certainly not, but certain phenomena of the crash might be. Avalanches, on the other hand, have always been very unpredictable events at least and seem to be quite good examples for chaotic behaviour. Summing up, we can say that dynamical laws – as understood in the usual way – describe physical systems which satisfy all four conditions D1, D2, D3 and D4. And this holds also for most applications of dynamical laws like CM, CEM and QM. However, the necessary conditions for dynamical laws are only D1 and D2; therefore, one of D3 and D4 (or both) may not be satisfied, and physical systems may obey D1 (D2) but not D3 or D4. It was a discovery of the late 20th century that chaotic motion, in the sense of "dynamical chaos", satisfies D1 (D2) but violates D3 and D4. Moreover, other kinds of motion, satisfying D1 (D2), do not satisfy D3 or D4 either, as the example of Maxwell shows.

7.2.3.3 Properties of Statistical Laws

- S1 The state of the physical system at t_i is not a definite function of an earlier state at t_{i-1}. The same initial state may lead to different successor states (branching).
- S2 Statistical laws describe and predict the states of the whole physical system, but they do not describe or predict the individual parts (objects) of this system.
- S3 Statistical laws describe only physical systems which are non-periodic, i.e. systems with extremely improbable recurrence of the whole state of the system.
- S4 The loss of information (and, consequently, the difficulty of prediction) about the state of an individual object (or a small part) of the whole system increases exponentially with the complexity of the system. On the

[36] The one which Maxwell refers to is "The same causes will always produce the same effects", which he discusses earlier. See below Sect. 9.4.4.

[37] Maxwell (1991, MaM), p. 13.

other hand: (accuracy of the) information about the average values of magnitudes (parameters) of the state of a huge number of individual objects (or particles) increases also with the complexity of the system.

7.2.3.4 Properties Compared: Reduction is not Possible

7.2.3.4.1 D1 Compared to S1

It is easy to see that there is an essential difference between the conditions D1 and S1. Like D1 is necessary for dynamical laws, S1 is necessary for statistical laws. This presupposes, however, that we interpret S1 (and, by it, statistical laws) realistically (i.e. in an ontic sense). That is, we assume there is real branching in reality. An epistemic interpretation according to which branching is only a sign for our lack of knowledge, whereas in the underlying reality everything is determined (by hidden parameters and dynamical laws of which we are ignorant), we do not find justified.[38] This can be substantiated by the fact that the following types of processes do not satisfy D1 (but satisfy S1), as is evident from all the sophisticated knowledge we possess today about these processes:

Thermodynamical processes, processes of friction, of diffusion, of radiation, of electric transport, processes of quantum mechanics, processes of biology, of cosmology, and of psychology.

If D1 is understood as the defining condition of determinism, then a system is deterministic if it satisfies D1. In this case, it follows that all processes which do not satisfy D1 are not deterministic in the sense of Laplace's determinism. However, if D1 only implies determinism (without being implied by it), then the fact that the above enumerated processes do not satisfy D1 does not imply anything about determinism. Thus, in this case it is possible that there are "hidden parameters" and unknown deterministic laws on "the ground" of the statistical behaviour on "the surface". This is, in fact, what Planck hoped, that the situation with the law of entropy could be as it is expressed in the quotation of 7.2.1.1 above. And later, similar interpretations with hidden parameters were proposed for the statistical laws of QM. If these were correct, one could hope for a theory which reduces statistical laws to dynamical laws. But, as will be shown in the comparison between D2 and S2, such a view is hardly tenable: Especially in QM there are certain experiments which seem to show unambiguously that the degrees of freedom allowed by statistical laws are real and that particles can be objectively undetermined. If this is correct, then the difference between D1 and S1 is sufficiently large such that neither statistical laws are reducible to dynamical laws nor vice versa.

On the other hand, it does not follow from this that these processes are non-causal, as will be shown in Chap. 9 below.

[38] cf. Weingartner (1998, SLG).

7.2.3.4.2 D2 Compared to S2

Similarly, D2 and S2 differ in an important point. Statistical laws are bound to huge ensembles; they describe physical systems consisting of a huge number of objects. The greater the number of objects, the more strict is the law about the whole ensemble. Though there is indeterminacy for every individual system, there is a strict law for the whole system if the ensemble is large enough. To some extent, such laws "emerge" from the "lawless" behaviour of a large number of individual systems. In this sense, Wheeler spoke of "law without law".[39] This problem was clearly understood and emphasised already by Boltzmann and Poincaré: How can the law of entropy emerge from random behaviour of individual systems? Schrödinger gave the following answer in his inaugural lecture of 1922:[40]

> "In a very large number of cases of totally different types, we have now succeeded in explaining the observed regularity as completely due to the tremendously large number of molecular processes that are cooperating. The individual process may, or may not, have its own strict regularity. In the observed regularity of the mass phenomenon the individual regularity (if any) need not be considered as a factor. On the contrary, it is completely effaced by averaging millions of single processes, the average values being the only things that are observable to us. The average values manifest their own purely *statistical regularity*."

This description fits very well to the statistical laws in thermodynamics. Concerning the statistical laws of quantum mechanics, Schrödinger's observation that the regularity of the single processes need not be thought to imply the laws of the ensemble can explicitly be demonstrated. Though it is clear also, here, that the theory refers to big ensembles of identically prepared systems, in quantum mechanics it may happen that the respective properties of individual quantum systems, like photons or electrons, are objectively undetermined. This can be illustrated in detail by means of so-called split-beam experiments, which deal with individual photons and other particles.[41] The main problem, here, is that for every individual system, say an individual photon, the value of the observable, before the measurement, is objectively undetermined, whereas a sufficiently large number of photons satisfy a statistical law, telling relative frequencies in the experiment. Thus, although there is indeterminacy in a good objective sense for every individual system, there is a strict law if the ensemble is large enough, such that we can speak of an objective and definite (i.e. Yes/No) property of the whole system. Despite of this particular situation in

[39] Wheeler (1983, RLL). For more details on this question see Sects. 12.2. and 12.3.
[40] At the ETH Zürich. This lecture was later published under the title "Was ist ein Naturgesetz?". cf. Schrödinger (1961, WNG), p. 11.
[41] For the emergence of statistical laws in quantum mechanics see Mittelstaedt (1997, ESL) and Sects. 12.2 and 12.3 below.

QM, it holds for all important statistical laws that the individual system is not definitely described by the law, but has its degrees of freedom which are not restricted by the law. Moreover, it has been emphasised that there are areas where not only individual particles but whole subsystems locally dodge the statistical law, in this case, the law of entropy: Every living system (organism) which produces and increases order, orthogenesis (maturation), information, and differentiation is a case in point. An example of a big subsystem of that sort is the earth embedded in the sun–earth–cosmic environment. (cf. Sect. 7.2.3.4.3 (2d) below). This shows unambiguously the difference between D2 and S2. And it shows again that, with respect to D2 and S2, neither type of law is reducible to the other.

7.2.3.4.3 D3 Compared to S3: Reversibility and Non-Recurrence

The difference between dynamical and statistical laws, which is usually viewed as the striking difference, is that which is expressed in D3 and S3: Dynamical laws are invariant under time reversal, statistical laws are not. The former describe processes which are time (reversal) symmetric, the latter describe processes which are irreversible;[42] or, more accurately, they describe processes with a very low probability of recurrence of the state of the system. It is also said that the statistical laws "define" an arrow of time or a time asymmetry, whereas the dynamical laws do not. Such formulations, however, have already been criticised above (6.4.2). Observe also that there are thermodynamic processes which are describable by statistical laws and which are reversible and recurrent as it is the case with Carnot processes. In order to show more accurately that these differences between the two types of laws prevent a reducibility of one type to the other, we first shall discuss time reversibility, and, secondly, time asymmetry, or better: non-recurrence.

(1) Time reversibility
    First of all, it should be mentioned that D3, i.e. the condition that the physical system (described by dynamical laws) is periodic, is not equivalent to saying that the system is time-reversal symmetric. Though periodicity – provided that the state *exactly* repeats itself after a finite period of time – implies time reversibility, the other implication does not hold. In the discussion of D3 above, it was mentioned that there are systems which obey dynamical laws but are not periodic, like systems of dynamical chaos. But these are, of course, non-recurrent but the underlying dynamical laws are time-reversal symmetric. In this respect, they are similar to systems described by statistical laws, but they differ because they satisfy D1 which is not satisfied by the latter. If Laplace's idea described by the illustration (recall the discussion of condition D1 above) would be a correct description of the developing universe, then time-reversal symmetry

[42] To mention immediately a terminological clarification, we shall not use the term "irreversible" or "irreversibility", but rather the more modest term of non-recurrence instead, for reason to become clear subsequently.

> would hold in the processes of the universe. But such a time development needs not to be periodic (even if subsystems like a planetary system could be). We know that time reversal is not the case in the universe; expansion and radiation suffice as examples.

It is a rather difficult question discussed among physicists whether the fundamental laws of physics can be time-reversal symmetric despite the many time asymmetries in the universe; in the sense that symmetrical laws together with asymmetrical conditions can describe asymmetrical phenomena (recall 5.4.2):

> "The disparity between the time symmetry of the fundamental laws of physics and the time asymmetries of the observed universe has been subject of fascination for physicists since the late 19th century."[43]

> "Next we mention a very interesting symmetry which is obviously false, i.e. *reversibility in time*. The physical laws apparently cannot be reversible in time, because, as we know, all obvious phenomena are irreversible on a large scale: 'The moving finger writes, and having writ, moves on.' So far as we can tell, this irreversibility is due to the very large number of particles involved, and if we could see the individual molecules, we would not be able to discern whether the machinery was working forwards or backwards."[44]

Keeping the laws (time-reversal) symmetric and putting the responsibility for the time asymmetric phenomena into the initial or boundary conditions, leads to explanations like the following:[45] the thermodynamic asymmetry presupposes progenitor states far from equilibrium; the CP asymmetry presupposes a spontaneous symmetry breaking of the Hamiltonian; the expansion of the universe presupposes a special singularity (big bang), etc.

However, many authors have also discussed time symmetric models of the universe.[46] In this case, the universe undergoes expansion and contraction in a symmetric way such that we have periodicity (even if the time of the sections might be long to allow the recurrence). However, there are at least two difficulties with such proposals: The improbable recurrence and the T-violation in weak interactions. The first is expressed by the following quotation:

> "The difficulty with the time-symmetric models is their implausibility. They require a very finely tuned set of boundary conditions, for which no explanation is offered. And amongst a number of other problematic features, it is difficult to imagine how a universe similar to the one in

[43] Gell-Mann, Hartle (1994, TSA), p. 311.
[44] Feynman (1997, SNP), p. 28.
[45] For the problem of the demarcation between laws and initial conditions see Chap. 8. Though, it should be remarked that there are many areas where this separation is not problematic.
[46] Gell-Mann, Hartle, ibid. Chap. 22.5. Schulman (1994, TSC).

> which we live, with large scale structures, stars and galaxies, might emerge, and then return to its initial state."[47]

The second was already briefly discussed in Sect. 5.4.3. Since CPT (charge–parity–time) invariance (of laws) is generally satisfied – all fundamental field equations are CPT invariant – but CP invariance is slightly violated in weak interactions, T has to outbalance the difference. Therefore, T invariance (time-reversal invariance) is not completely satisfied for the fundamental laws of nature.[48] CPT invariance – one of the most important symmetries of quantum field theory – says that physical laws seem to be symmetric with respect to the complex exchange of particle–antiparticle, right–left, and past–future. This CPT symmetry has remarkable consequences: It implies that the mass of the (any) particle must be the same as that of its respective antiparticle. The same holds for their lifetimes. Their electric charges must have the same magnitude but opposite signs, and their magnetic moments must agree.

A most striking consequence, however, is that, since CP invariance is violated but CPT invariance is not, T has to outbalance the difference, and therefore, T invariance cannot hold unrestrictedly. This is, indeed, a serious consequence and – if true – would have a lot of implications. On the other hand, T-invariance (time reversal) is hardly compatible with a dipole-moment of elementary particles, and it would reverse velocities and exchange initial with final states. Such a time reversal operation would be strongly non-linear in character, as Wigner pointed out long ago. If it could be proved that the neutron has a dipole momentum, then the laws of nature could neither be P-invariant nor T-invariant. As it has been said above, we know from other experiments that they are not P-invariant. Since CP invariance is also violated but CPT invariance is not, it is important to ask for the basic assumptions which underlie the CPT invariance. These are mainly the following:[49]

(1) The particles which are not composite are finite in number.
(2) The symmetries of special relativity hold (with respect to continuous transformations and one time and three space coordinates).
(3) The laws are local.
(4) Energies cannot be arbitrarily negative, i.e. there is a lowest energy level.
(5) The laws of QM hold in accordance with local relativistic quantum field theories in four dimensions.
(6) The total probability of the quantum system is constant in time.

Moreover, as it appears from recent experiments, T-reversal symmetry seems to be violated directly, too, and not only via CPT symmetry and CP violation.

---

[47] Halliwell (1994, QCT), p. 374.

[48] This holds, provided there are no other ways out; for instance, that the T-violation is not due to an asymmetry in the cosmological boundary conditions or to an asymmetry of our particular epoch and spatial location. cf. Gell-Mann, Hartle (1994, TSA), p. 329.

[49] cf. Wess (1989, CPT) and Genz, Decker (1991, SSB), p. 169.

The violation concerns weak interactions. But, since weak interaction concerns all elementary particles except photons, the experimental result appears to be very important. There have been two different series of experiments independently made in CERN and FERMILAB, which seem to prove the violation: The time dependent rates for the strangeness-oscillation process $K^0 \to \bar{K}^0$ and its inverse, $\bar{K}^0 \to K^0$ (neutral kaons) are different. If time-reversal symmetry were strictly preserved, we should have identical rates (CERN).[50] The second experiment was based on the following idea. Both the time-reversal operator T and the parity operator P reverse the direction of the momentum of a particle (T reverses, in addition, the particle's spin). The experimentalists measured for each observed decay the angular variable ϕ that changes sign when all the final state momenta have their directions reversed. Time reversal symmetry would require that the observed ϕ and sin 2ϕ distribution be symmetrical about zero. The observed asymmetry is about 14%, which is in agreement with the theoretical expectation.[51]

(2) Non-recurrence
   (a) Examples for recurrence and non-recurrence
   Skiing in fresh powder snow is a great pleasure. But if the slope is small and one is skiing down frequently, the slope will be filled with traces, and after some time no new space (powder snow) is left, and thus one has to use one's own traces again (recurrence). This illustration tells us already some important conditions: The motion has to be area preserving (the skier is not supposed to leave the slope) and in a finite region. Observe now that just by raising the complexity of the system, recurrence becomes very improbable: Imagine that there are thousands of skiers on the slopes of a big ski resort (the cable cars and lifts of a big ski region in Austria can take up about 60,000 people per hour). The probability that at some later time t_1 all skiers will be again at a position in which all skiers were at the time t_0 earlier such that the whole state of this system would recur has much lower probability (even if we assume that they go on skiing days and nights) than the recurrence of a single skier to an earlier position.

   But observe also the following: Living organisms act randomly for some time but not always. And even if they act according to certain goals, there are two possibilities: The goals are different, and then the recurrence of the whole system (encompassing a huge population) will still be very improbable. The second possibility, however, is that the goals agree with respect to one or more properties: Thus, all skiers will have dinner in restaurants and moreover will return to their guest rooms in hotels and pensions for the night; i.e. with respect to the sleeping places, we will have recurrence at the second night and so on (assuming that the skiers stay, say a week, in that ski region).

[50] Angelopoulos, et al. (1998, CPL).
[51] Schwarzschild (1999, TEO).

(b) Boltzmann's example
Instead of living organisms take the molecules in a litre of gas (air) at temperature $T = 0°\mathrm{C}$ (273 K) and atmospheric pressure ($1.033 \cdot 10^3$ g cm^{-2}). Of these, we shall not assume that they "act according to certain goals". A litre of air (at temperature and pressure mentioned) consists of $2.688 \cdot 10^{22}$ molecules. It will be understandable that this system of $2.688 \cdot 10^{22}$ molecules can be in a huge number of different (micro-) states. The number is about $10^{5\times10^{22}}$, so as to realise the macrostate "litre of air under the conditions mentioned". Thus, the same (for our eyes or lungs the same) macrostate can be realised by a huge number of different microstates. Boltzmann's discovery was that the probability of such a macrostate can be defined as the number of microstates which can realise the macrostate, and that this number (more accurately the logarithm of it) is the entropy. The probability that such a macrostate of a physical system occurs by chance, i.e. out of a huge number of microstates, describes the degree of disorder (the entropy) of this system. Assume we have two macrostates with probabilities p_1 and p_2, then the probability that they both occur will be $p = p_1 \cdot p_2$. The associated entropies S_1 and S_2, since they are additive, will lead to $S = S_1 + S_2$. In order to connect probability p with entropy S, it tells from mathematics that the relation must be logarithmic, with some constant, which appears to be Boltzmann's constant. Thus, we get Boltzmann's famous law: $S = k\, ln\, p$.

What is the probability of the recurrence of one of the microstates in a litre of air? It is 1 in $10^{5\times10^{22}}$. Assume one litre of air expands into a 2-litre vessel. What is the probability of the recurrence of the former state, i.e. that it will occupy only 1 litre again? For each molecule the probability is 1/2; thus, for two molecules the probability is $(1/2)^2$; for three: $(1/2)^3$, etc. and for n molecules $(1/2)^n$. Since $n = 2.7 \cdot 10^{22}$, it will be understood that the probability is extremely low.

We might ask the questions: Will all the $10^{5\times10^{22}}$ microstates of the litre of air be realised at all? And in what time? This leads to an interesting cosmological question: Assume that we are asking how many possible microstates are in the whole universe, in order to calculate the entropy of the whole universe. Then, the question arises whether every microstate can be realised within the lifetime of the universe, if the lifetime is finite. Since the number of microstates is extremely huge, they probably will not all be realisable within the lifetime calculated by the standard (big bang) theory. If this is so, then there are more possible universes than the actual universe, which obey the same laws of nature and differ only with respect to some microstates from each other. In other words, the laws of nature have more (possible) models than the one actually realised.[52]

[52] For more on that see Chap. 8.1.6.

(c) Poincaré's recurrence theorem

Poincaré's recurrence theorem says that for a (time independent) Hamiltonian system a trajectory returns to a given neighbourhood of a point in a sufficiently long (or in an infinite) time. Zermelo thought that he had proved that this theorem shows that Boltzmann's statistical mechanics cannot be a correct description of thermodynamic processes as irreversible processes or processes with extremely improbable recurrence.[53] However, as the replies of Boltzman[54] show, Zermelo partially neglected and partially misunderstood important conditions in connection with Poincaré's recurrence theorem. The first thing which is made clear by Boltzmann is that Poincaré's recurrence theorem is not applicable under conditions where the number of molecules is infinite and time is increasing and can be very long but finite. On the other hand, if the conditions are such that time is infinite and the number of molecules is very large (but finite and in a finite space), then Poincaré's theorem is applicable.[55] And eventually, if time is finite but very long and the number of molecules is very large – and this is the realistic situation – then there is no incompatibility, since Boltzmann's principle claims only that the recurrence is extremely improbable. The second thing, which Boltzmann realised very clearly, is that the probability of recurrence depends very much on the complexity (for example with the number of molecules, as shown above). In fact, Boltzmann was modest enough to underline the enormous probability (for non-recurrence). And if he uses irreversibility at all, he adds immediately that what is at stake are very extreme probabilities.

A further point which is worth to be mentioned is that Boltzmann's entropyis not identical with Gibb's entropy.[56] Whereas the first is defined for a microstate of a macroscopic system, the second is defined for an ensemble density of a microstate. They agree for systems in local equilibrium. But they differ in their time development: The Gibb's entropy does not change in time when the ensemble density develops in time, while the Boltzmann entropy does. At the starting time of a system, in the state of local equilibrium, both entropies agree, but subsequently Boltzmann's entropy would increase, while Gibb's would not change. Therefore, the important entropy for describing the time development of macrosystems is the entropy of Boltzmann.

[53] Zermelo (1896, SDM) and (1896, MEI).
[54] Boltzmann (1896, EWB) and (1897, ZAM).
[55] But in an infinite time an earlier state can always recur. cf. Boltzmann (1896, EWB), p. 569.
[56] cf. Lebowitz (1994, TAB).

(d) Irreversibility locally violated

Erwin Schrödinger raised the question: How can we understand a living system (living organism) in terms of Boltzmann's theory?[57] Or, how can these systems manage to keep, or even to increase, a low entropy level despite of the validity of the law of entropy?

Or, again, in other words: How are living systems capable to decrease the disorder and the probability of realising their macrostates on the one hand, and to increase order, to create new subsystems and raise the level of information, on the other? The answer, which was partially already given by Schrödinger, includes the following points:

(i) Living systems (organisms) are not thermodynamically closed; they are open systems.
(ii) Living systems receive high grade energy (energy with low entropy) from their environment via metabolism, but they pass on low grade energy (energy with high entropy).
(iii) By process (ii) the living systems are capable of achieving orthogenesis (maturation), i.e. increasing order, quality, and differentiation.

Cosmological investigations show that also planets, especially the earth, behave in a similar way as living systems. For example, the earth satisfies the above three conditions in an analogous way: It receives high grade energy as electromagnetic radiation (with Planck temperature of 5600 kelvin) from the sun and passes on low grade energy as heat radiation (with Planck temperature of only about 300 kelvin) into its environment. The received energy has low entropy and a respectibly lower number of degrees of freedom(i.e. of possible microstates which can realise a macrostate), whereas the delivered energy has high entropy and a much greater number of degrees of freedom. From the above consideration about living systems, it is also understandable that the earth *with* living organisms (including men) produces more entropy than it would produce without. It should be mentioned, however, that, although the flow of entropy is not balanced, the flow of energy (received per time unit and delivered per time unit) is balanced. That means that the earth, and also all living systems, convert high grade energy, by passing it through their systems, into low grade energy, and by that process they are able to create order, quality, and differentiation. The Earth (and living systems) are open systems of non-equilibrium, which permit a loss of entropy which does not violate the law of entropy, since in the whole system sun–earth–cosmic environment, entropy still increases.[58] But it shows that, locally, the lawlike direction of the thermodynamic processes (towards higher entropy) can be reversed such that order and information can be produced.

[57] Schrödinger (1944, WLf).
[58] cf. Fahr (1997, WKS).

(e) Non-recurrence versus irreversibility
From the above considerations (a)–(d), it will be understandable that we want to avoid the term "irreversibility" for three reasons:
   (i) What thermodynamic processes and many others, like radiation, cosmological expansion, processes of measurement, biological and psychological processes, really show is that recurrence of the state of the whole system is very improbable, but not that recurrence or time reversal is impossible. The latter cannot be proved by a statistical law. Thus, we claim only what can be empirically corroborated with the help of statistical laws.
   (ii) From the last example described in (d), it is plain that thermodynamic processes can be reversed locally without violating the second law.
   (iii) Moreover, we have seen above 7.2.3.2 that non-recurrence and time-irreversibility are not equivalent notions: since we have cases of non-recurrent phase density and time reversibility in chaotic motion of dynamical chaos.

Summing up. A comparison of D3 and S3 shows that the difference between dynamical and statistical laws, which is usually viewed as the most striking one – time reversal invariance of the laws versus irreversibility of the laws – has to be taken with care. Strict time-reversal symmetry is not any more valid on the microlevel, and irreversibility on the macrolevel should be better replaced by very improbable recurrence. However, the differences in this respect are sufficiently large to forbid reducibility in the one or the other way. But the more careful interpretation paves the way for the compatibility of both types of laws.

7.2.3.4.4 D4 Compared to S4

From what has been said under Properties of Dynamical Laws (Sect. 7.2.3.2), it will be clear that a physical system which obeys D4 cannot be a chaotic system, because the violation of D4, i.e. sensitive dependence on initial conditions or exponential separation of adjacent conjugate points (with respect to the starting point), is a necessary condition for dynamical chaotic motion. Another necessary condition for dynamical chaotic motion is the average loss of information about the position of a point in an interval (relative to one iteration). Both the separation of adjacent points and the loss of information about their positions is measured by a positive Lyapunov exponent. From this, it follows that the presence of D4 (stability) preserves the system from both, sensitive dependence on initial conditions and average loss of information. Therefore, it is the more understandable that in the presence of D1 and D4 also D2 can be satisfied.

On the other hand, statistical laws are characterised by a loss of information (and prediction) about the state of an individual object (a point, a small part), (S4). And this loss of information increases exponentially with

the complexity of the system. In this point, therefore, there is at first sight a similarity with the behaviour in chaotic motion. But the difference is that the indetermination (degrees of freedom) for the individual case and the loss of information about it is an essential property of the description of the system by statistical laws, whereas in the case of dynamical chaos it is an exception, since by the underlying dynamical laws also every individual subsystem should be describable by dynamical laws (D2). However, in the case of statistical laws, the accuracy of the information about the average values (think of velocity distribution of particles, etc.) increases with the complexity of the system. There does not seem to be a parallel of this in dynamical laws, independently whether they obey all D1–D4 or only a part of it (like the underlying laws in dynamical chaotic motion). Moreover, chaotic motion of dynamical chaos is not in general describable by statistical laws, i.e. there is no general predictability of average values.

Therefore, the result of the above consideration is this: Condition D4, which is also called the condition of *stability* or the condition of *robustness*, preserves the dynamical properties of the system; i.e. it preserves the conditions D1 and D2 (and for many systems D3). Since S4 differs essentially from D4 with respect to the description of the individual case, no reduction of statistical laws to dynamical ones or vice versa is possible. Moreover, S4 differs with respect to the average values of magnitudes also from the conditions for dynamical chaos (where D4 is not satisfied), which shows another aspect of the non-reducibility of the two types of laws.

7.2.3.5 Compatibility of Dynamical and Statistical Laws

Concerning the compatibility of dynamical and statistical laws, we shall deal with two points: (1) With an illustration to show that both types of laws can be compatible even within the same physical system. (2) By showing that the so-called arrow of time is not a hindrance for the compatibility.

(1) First of all, it has to be remembered that basic laws which are time symmetric can "produce" asymmetric phenomena (states and processes) if there are asymmetric initial conditions (cf. Chap. 5. above); i.e. an asymmetrical world with asymmetrical states and processes does not imply asymmetrical laws but requires asymmetrical initial conditions. Asymmetrical initial conditions of a very strong kind, in fact the greatest symmetry breaking – also with respect to time – have to be assumed at the beginning of the universe in all the theories which describe a universe finite in time or with a finite age since the beginning. This amounts to an extreme and – in the sense of the law of entropy – most improbable but most ordered and structured singularity.

    Secondly, it should be emphasised that basic laws which are time symmetric on the microscopic level are compatible with laws on the macroscopic level which are not time symmetric, but describe an arrow of time

like the law of entropy (even if granted that the reverse process is not completely impossible but very highly improbable). A nice "Gedankenexperiment" for such a compatibility is given by Lee.[59]

Assume a number of airports with flight connections in such a way that between any two of these airports the number of flights going both ways along any route is the same. This property will stand for microscopic reversibility. Some of the airports may have more than one air connection (they are connected with more than one other airport), whereas other airports have a connection only to one airport (let's call such airports dead end airports). A passenger starting from a dead end airport (or starting from any other airport) can reach any other airport and can also get back to his starting airport with the same ease. This property stands for macroscopic reversibility. In this case, we have both microscopic and macroscopic reversibility.

But suppose now we were to remove in every airport all the signs and flight information, while maintaining exactly the same number of flights. A passenger starting from a dead end airport A will certainly reach the next airport B, since that is the only airport connected with A. But then – especially when assuming that B has many flight connections – it will be very difficult to get further to his final destination; in fact, it will be a matter of chance. Moreover, his chance to find back to his dead end airport A will be very small indeed. Moreover, if millions of flight passengers fly around without any flight information, the probability of recurrence of the whole system (say that each passenger is again in an earlier position at some specific time t) will be extremely improbable.

Thus, in this case we have microscopic reversibility maintained but macroscopic irreversibility, and both are not in conflict.

(2) Is the "arrow of time" compatible with dynamical laws?
We shall discuss this question with respect to three different descriptions of time which say that: (a) time flows, (b) time flows only in one direction, (c) time is connected with directional processes.
(a) Time flows
That time *flows* we grasp from change, i.e. mutation and movement (i), with respect to an ordered sequence (ii).
(i) Without any change (mutation or movement) time would "stand still" such that change (mutation and movement) is a necessary condition of time – at least for our understanding: "It is utterly beyond our power to measure the changes of things by time. Quite the contrary, time is an abstraction at which we arrive by means of the changes of things."[60]

[59] Lee (1988, SAW).
[60] Mach (1960, MEC), p. 273. Recall also the definitions of time given by Aristotle and Thomas Aquinas in Sect. 6.2.2.

That time presupposes change was already pointed out by Aristotle in his definition of time (recall Sect. 6.2.2(1)). In a different context – the transcendental philosophy – the same point was made by Kant in the *Critique of Pure Reason.* Since time cannot be observed, the direction of time is not given by experience such that the temporal order of events could be determined by reading off the respective time values from a fictitious universal clock. By contrast, causally connected events determine the order of instants of time such that the cause event is always earlier than the effect event. It is obvious that in this way a *metric* of time cannot be established but merely a *topology.*[61]

That there is no past to future direction of time in regions that are at equilibrium was pointed out by Boltzmann.[62] He compared this with gravitation: As there is no downward direction in regions of space where there is no gravitational force, there is also no past to future direction of time in regions that are at equilibrium. Boltzmann's claim is stronger than Aristotle's, if "equilibrium" is understood as thermodynamic equilibrium. But, if we understand "equilibrium" in a wide sense just as "no change whatsoever", then Boltzmann's point is the same as Aristotle's: time presupposes (some kind of) change. This change can be of different kind, but it will always belong to one of the two following groups: Change in the physical and biological sense: Any movement of material bodies (in the universe) or any change in living organisms. Change in the psychological or mental sense: Any conscious proceeding of thinking, feeling and desiring. That the understanding of time is also supplied by such "inner" mental experiences is often forgotten. If it is underlined here, it is not claimed that this kind of understanding of time is independent of the one based on experience with the "outside world", but merely, that it plays also an important additional rôle for our understanding of time.[63]

(ii) Thus, it seems better to speak of the asymmetry of a flowing process of a sequence of successive states that are ordered by a partial ordering, instead of a "flowing time". In such a sequence, we distinguish past and future states, and we measure the distance between them with the help of time units (produced by another physical periodic process in a clock). So far, no direction is presupposed; only partial ordering and the distinction between

[61] Kant (1787, KRV), B233.

[62] Boltzmann (1897, ZAM), p. 583.

[63] This psychic or mental aspect in our experience of time has been stressed by different authors. cf. for example Augustin (Conf) XI, 27 and 28, and Eddington (1928, NPW), p. 91ff. cf. further Sklar (1977, CTT).

past and future; i.e. intervals and units. The successive time units (first, second, etc.) which correspond to the successive states may have suggested the idea that “time flows”.

The presupposed ordered sequence is best described by the chronology of time or by the basic axioms of tense logic, which assume partial ordering, transitivity, asymmetry, irreflexibility and density (recall Sect. 6.3.3). In such a sequence there is already the distinction between earlier and later or past and future, because this distinction is understood either only in analogy to – or more usually as a mapping to – the distinction of smaller and greater in the sequence of (real) numbers. Therefore, it is not correct to say that understanding of the distinction between earlier and later (or past and future) presupposes (understanding of) directional processes, although it is correct that it presupposes (understanding of) some kind of change which can be also non-directional, like periodic change.

Expressed with the above idea that time flows, that means (α) that “time flows equably” everywhere (in the whole universe). In the second interpretation (β), however, this assumption is uncovered and then avoided, because the dependence of the measuring units (time units) on the corresponding successive states of real processes are taken into account. Consequently, the measuring units (time units) are not understood to be applicable universally as rigid. Expressed with the above idea, that means that “time flows unequably” in general, although it may “flow equably” locally.

The first interpretation (α) is that of Newton:

> “Absolute, true and mathematical time, of itself, and from its own nature, flows equably without relation to anything external”[64]

The second interpretation (β) is that of Einstein in GR:

We may paraphrase it in a way, analogous to the text of Newton, as follows: The relative time of an observer plus reference frame is measured by standard clocks. By stipulation it flows equably but it depends on the distribution of matter, fields, and boundary conditions. Every observer plus reference frame has its own time scale, and there is no universal time scale that is relevant for all observers. The continuous time translation or the time development expressed by dynamical laws is in accordance with the first interpretation under the condition of a classical approximation and restriction; it is in accordance with the second interpretation under the condition of a generalisation and relativisation with respect to the whole universe.

[64] Newton (Princ) I, Scholium. Recall Sect. 6.2.2.(3).

Summing up: This point (a) (i.e. that "time flows") which implies change and partial ordering and intervals measured in units is, of course, presupposed by dynamical laws, and so there cannot be any incompatibility.

(b) Time flows only in one direction

Concerning (a) ("time flows)" we cannot speak of an "arrow" of time. But an arrow of time (of different strength) seems to be expressed by (b) and (c). Observe further that (c) includes (b) and (b) includes (a).

However, the subsequent considerations in (b) and (c) will show that the frequently used expression of the "arrow of time" is misleading and not justified: although "time flows" in the sense of passing away, the arrow is intrinsic *in processes* (not "in" time). That time "flows only in one direction" we grasp in a twofold way: first, naturally and scientifically, by our experience, with an ordered sequence of particular changes in a process. This first way will be considered in (c), whereas the following second way will concern us here. Second, scientifically, by comparing a particle's movement on a spatial coordinate with a particle's movement on the time coordinate. Concerning this question of how to distinguish a (particle's) movement on a spatial coordinate (in GR: space like geodesic) from a (particle's) movement on a time coordinate (in GR: either time like geodesic or null geodesic[65]), we may formulate two subquestions: (α) Can the coordinates (geodesics) be distinguished by their directions (vectors)? (β) Can the coordinates (geodesics) be distinguished by their closure conditions? Both subquestions can be answered with: Yes. The answer to the first subquestion is very well expressed by the following quotation from Wigner: "The difference between the two cases arises from the fact that a particle's world line can cross the t = constant line only in one direction (in the direction of increasing t); it can cross the x = constant line in both directions. If we replace "line" in the last sentence by "plane", we have the generalisation of the distinction to the actual four-dimensional universe."[66]

The answer to the second subquestion is the following: According to GR, the space of the universe is closed (even if the universe is expanding); that is, there are closed spatial coordinates or closed space-like geodesics. On the other hand, we usually assume that the time coordinate is not closed; i.e. we assume that the non-space-like geodesics (time-like geodesics and null geodesics) are not closed. This assumption has been called the *chronology condition* of spacetime.[67] This condition plays an important role for the concept of causality. Causality would break down and one could travel into one's own past,

[65] For space like, time like and null geodesics see Hawking, Ellis (1973, LSS).

[66] Wigner (1972, TEU), p. 239.

[67] cf. Hawking, Ellis (1973, LSS), p. 189.

if the chronology condition is not satisfied (cf. Sect. 9.2.2.1 on causality). This second point (b), that the time coordinate is distinguished from the spatial coordinates such that a particle's world line can cross the t = constant line only in the positive direction of increasing t, is not determined by dynamical laws; because dynamical laws permit both directions positive and negative. From this, it follows that (b), i.e. *time flows in one direction*, is compatible with dynamical laws, because dynamical laws allow both directions.

(c) Time is connected with directional processes
In addition to the directional feature of time described in (b), which is not at all in conflict with dynamical laws, there are two further features of time which are connected with directional processes. A directional process can appear in a twofold way: First (i), through the behaviour of huge ensembles of objects in contradistinction to the behaviour of singular objects. Second (ii), when (i) is accompanied by the increase of a certain physical magnitude.

(i) Consider the following example: Assume a very large (long) but finite sequence of decimal places after 0, say the (finite) sequence of natural numbers, i.e. 0, 1, 2, 3, ..., 10, 11, 12, ..., 99, 100, 101, ..., etc. It can be proved that this sequence has a normal distribution. It will, therefore, be easily understandable that the probability of recurrence for the three numbers 1, 4, 2 (in this order) on decimal places will be not very low. It occurs in 142, in 1420, 1421, etc. In contradistinction to that, the probability of the recurrence of an ordered sequence of 10^{10} numbers on decimal places, as part of the above sequence, will be very much lower. This has nothing to do with entropy or with the increasing of a certain physical magnitude. But it has to do with "direction" and asymmetry. The more the sequence of numbers increases, the more improbable is the recurrence of the (whole) sequence. This shows that for such an asymmetry (or "direction") no physical process is needed, because the objects in the huge ensemble may be conceptual objects. But let the objects be, now, atoms of a fluid or gas; and instead of a sequence, we may have a structure, i.e. a state or a distribution in a phase space. It will be clear just by probability considerations that the recurrence of a state with 10^{10} atoms is much much lower than that of a state with three atoms.[68]

Concerning the question of the compatibility with dynamical laws, it should be clear that there cannot be any conflict of dynamical laws with an asymmetry or direction of a purely mathematical kind.

[68] Recall the examples in the section on Non-Recurrence above 7.2.3.4, D3 compared to S3 (2).

(ii) Directional processes: First, it should be clear that chronological time scales, as they are used for time measurement, do not define a direction of time even if they indicate that "time flows" in the sense of representing a sequence of partial ordering. Remember Poincaré's considerations concerning convention in this respect (Sect. 6.2.1.5) above. Assuming events (states) S_1, S_2, $\ldots (\in S)$, reference frames plus observer, RF, and chronometrical scales CS (mappings of durations onto real numbers, standard clock), the following postulates are basic for time T and time interval t:

(α) T is a function $\{S_1, S_2, RF, CS\} \rightarrow \Re^+$ with $T\{S_1, S_1, RF, CS\} = 0$.

(β) For every state S_1 relative to RF and CS and for any value $t \in \Re^+$ there exists a second state S_2 such that $T(S_1, S_2, RF, CS) = t$.

(γ) Transitivity. For any triple of states (S_1, S_2, S_3) relative to RF and CS it holds $T(S_1, S_2, RF, CS) + T(S_2, S_3, RF, CS) = T(S_1, S_3, RF, CS)$.

From these postulates, it follows that for any two events S_1S_2 relative to RF and CS it holds that: $T(S_1S_2RFCS) = -T(S_2S_1RFCS)$. This shows clearly that the above basic assumptions and postulates do not define a direction "in" time. They tell us only that the direction from S_1 to S_2 is the opposite of the direction from S_2 to S_1, but not which event is first in nature. Moreover, these postulates are in full agreement with the above mentioned paraphrase of Newton's statement about absolute time, if the reference system RF is understood as a reference frame plus observer and the chronological scale CS as a time scale of a standard clock.

On the other hand, so-called directional processes (of nature) tell us unambiguously which event is first and which is second. Penrose lists seven such directional processes:[69]

(1) The decay of neutral K mesons in weak interactions (recall Sect. 7.2.3.4.3(1) above).
(2) The process of measurement in quantum mechanics, especially the so-called "collapse of the wave function".
(3) All processes in which entropy increases (recall 7.2.3.4.3(2)).
(4) All processes of radiation.
(5) All conscious mental processes (recall 7.2.3.5(2a) above).
(6) The process of expansion of the universe.
(7) The process of the gravitational collapse ending in a *black hole*.

Of these processes, (1), (3), (4) and (6) are experimentally very well confirmed. (5) is very well confirmed by introspection and by the descriptions of the psychology of mental processes. The claim that (2) is a directional process

[69] Penrose (1979, STA).

is – at least to a considerable extent – a matter of interpretation of the quantum mechanical process of measurement. The time reversal of (7), leading to a *white whole* (no experimental evidence so far), is an open question such that (7) cannot be viewed as an unambiguous case of a directional process. Furthermore, it is an open question whether processes (1), (4), (6) and perhaps (7) can be reduced ultimately to process (3).[70] In this case, the underlying fundamental law would be the second law of thermodynamics, i.e. the law of entropy. Concerning (4), we refer to Chap. 8 below where the ambiguity of the solutions of Maxwell's equations will be discussed. With respect to (6) we have to say that it is well confirmed for the past 20 billion years, but there is, of course, no security for the long time future.

Now, concerning the question of a possible incompatibility between directional processes and dynamical laws, the following observations are important: A first observation is that such processes can be described and explained only by statistical laws and not by dynamical laws. A second observation is that both laws, and, therefore, also their respective processes are compatible in the same physical system as is shown by Lee's Gedankenexperiment (recall 7.2.3.5(1) above). A third observation is that *time reversal* of dynamical laws should not be taken too strictly, since it has been softened up already on the microlevel with respect to weak interaction. A fourth observation is that also *irreversibility* of directional processes should not be taken absolutely, because what is empirically defensible, is rather extremely low probability of recurrence.

7.2.4 Answer to the Objections

7.2.4.1 (to 7.2.1.1) A reduction of statistical laws to dynamical ones and, accordingly, a reduction of unidirectional phenomena to "conservative effects" (which permit time reversal) seemed to be possible, when Planck expressed his hopes in the quotation in 7.2.1.1. At that time, he was supported by Zermelo who was his assistant, and there was the controversy of Zermelo with Boltzmann in the Wiedemann's Annalen 1896 and 1897 (recall Sect. 7.2.3.4.3(2c)). But, after Boltzmann showed in his replies that Zermelo partially neglected and partially misunderstood important *physical* conditions concerning the restricted application of Poincaré's recurrence theorem, Zermelo seems to have left the subject and to become engaged very successfully in set theory.

Independent of historical matters, it has been shown in the sections of Sect. 7.2.3.4 – by a comparison of four characteristics of dynamical and statistical laws – that no reduction from one type of law to the other is possible. This was also supported by recent research concerning the properties of dynamical and statistical laws.

[70] cf. Wheeler (1994, TTd). Concerning (7), the important question is whether the horizon area of the black hole can be proved to be proportional to measures of entropy, which has been supported by Christodoulon, Bekenstein and Hawking.

7.2.4.2 (to 7.2.1.2) This objection makes two important assumptions: (1) That statistical laws express a lack of knowledge of the details of a state of a physical system. (2) That, if all details of a certain state are known, all past and future states can be calculated (according to Laplace). Concerning the first assumption (1), there is a right and a wrong interpretation. The right interpretation of (1) is that every statistical law implies some deficiency of information about the individual case, which increases with the size and complexity of the ensemble. The wrong interpretation of (1) is that all the individual details (of which there is lack of knowledge) are in fact fixed by hidden parameters or by a deterministic structure. That such an assumption is hardly tenable was shown in Sect. 7.2.3.4.1. What processes of thermodynamics, of friction, of diffusion, of radiation, of cosmology and of quantum mechanics show, is rather that there are real degrees of freedom, or there is real branching in nature.

This is connected with the second assumption (2): If all details of a certain state are known, then all past and future states can be calculated with the help of (dynamical) laws. This assumption is only correct if the conditions D1, D2 and D4 for physical processes obeying dynamical laws are satisfied. But, as we know, these conditions are not satisfied by the above mentioned processes, which cannot, therefore, be described by dynamical laws. Thus, (2) is not true in general.

Moreover, the antecedent of assumption (2) (i.e. "if all details of a certain state are known") is not always satisfiable, even if dynamical laws can be applied as in the case of dynamical chaos. In this case, then, assumption (2) is applicable only in an empty sense, or better, not applicable at all.

A further restriction on (2) – mentioned already in note 14 – is that Laplace's intelligence must not belong (as a part) to the physical system, and in case the physical system is the whole universe, the intelligence has to be non-material and not belonging to the universe.

Since the objection (7.2.1.2) uses the wrong interpretation of assumption (1) and makes assumption (2), which is not generally true, the conclusion in the argument (of the objection) is not proved.

8

Laws, Boundary Conditions, and Constants of Nature

> "The world is very complicated and it is clearly impossible for the human mind to understand it completely. Man has therefore devised an artifice which permits the complicated nature of the world to be blamed on something in which simple laws can be found. The complications are called initial conditions, the domains of regularities, laws of nature..." (E. P. Wigner)[1]

8.1 Are Boundary Conditions Independent of Laws of Nature?

8.1.1 Arguments Contra and Pro

(1) What is ruled by a common coherent cause is not independent from each other. But according to Mach's principle the matter distributed in the universe is the common coherent cause for both the laws and the initial conditions. Therefore – if Mach's principle is correct – then the initial conditions are not independent of the laws of nature.[2]

(2) If our universe is the only model (realisation) to satisfy the laws of nature, then the boundary conditions might be fixed by this model and thus be dependent on the laws of nature. But it seems that there is only one (this) universe.[3] Therefore, – on the assumption that there is only one universe – the boundary conditions are dependent on the laws of nature.

(3) A law of nature is something, which does not change, i.e. which is invariant (symmetric) relative to something, which changes (Chap. 5). Among those changing conditions the boundary conditions and especially the initial conditions are an important case in point, i.e. all laws of nature are

[1] Wigner (1967, SRf)), p. 3.
[2] It is not stated here that Mach's principle actually holds.
[3] The opposite point of view is defended, e.g. by Vilenkin (1982, CUN).

invariant under a change of initial conditions. But this is only possible if the initial conditions are independent of the laws of nature. Therefore, the initial conditions must be independent of the laws of nature.

8.1.2 The Problem of the Separation of Boundary Conditions and Constants from Laws of Nature

In the physical reality we observe various objects and processes, motions of material bodies, optical phenomena, and heat processes, but we do never observe directly laws of nature, boundary conditions, and constants of nature. However, it is a methodological aspect of our way to describe the reality in terms of physics, to splitting the phenomena into three distinct parts, the laws, the boundary conditions, in particular the initial conditions, and some constants of nature like gravitational constant or velocity of light. This way of description by separation is by no means new since it was already successfully applied by Newton in his mechanics. Whether this three-fold way to grasp the reality is useful or necessary in all cases is a difficult question, which will be treated in the subsequent sections. Some arguments in favour of this separation can also be found in Sect. 11.2.3, where we will discuss the objection that the laws of nature do not describe facts.

That the question does not have a simple answer can also be seen from the following consideration. What we understand by initial conditions is dependent to some extend on the question which point of view we find more natural at least in connection with observation. The more natural view for us is to understand initial conditions as characterising the state of the system (with all spatial coordinates) at a definite instant of time and then considering the development of the system into a state at a definite later instant of time. By contrast, the more unnatural view for us is to understand initial conditions as characterising the state of the system for all times, but only for a single instant of one of the spatial coordinates and then to formulate the spatial derivative according to this point of view.[4] There is of course a good reason for the preference of the former "natural" point of view: By empirical observations we are able to grasp very small and very huge distances in space (at one point of time) but only short distances in time (for instance the life time of a definite state). Therefore we use the *position–momentum* characterisation of a physical system rather than the *time- energy* characterisation.[5]

[4] cf. Wigner (1972, TEU) p. 239.

[5] Observe, however, that in Hamiltonian mechanics we could use in principle any pair of canonically conjugate observables for characterising the state of a physical system.

8.1.3 The Separation of Boundary Conditions and Constants from Laws of Nature is Possible

That this is so can be shown first by the elementary example of planetary motion, which also holds as representing the area of classical mechanics.

Example Planetary Motion

Two bodies with masses m_1 and m_2 and positions $\vec{r}_1$ and $\vec{r}_2$ are coupled by gravitational forces given by the potential $V(r) = -G\frac{m_1 m_2}{r} \equiv -\frac{\alpha}{r}$ with $r = \left|\vec{r}_1 - \vec{r}_2\right|$ and the gravitational constant G. Hence in the relative coordinate $\vec{r} = \vec{r}_1 - \vec{r}_2$ the equation of motion reads

$$\frac{m_1 m_2}{m_1 + m_2}\ddot{\vec{r}} = -\frac{\alpha}{r^3}\,\vec{r}\quad . \tag{1}$$

Since this equation is invariant against rotations the angular momentum is a constant of motion and the problem can be reduced to a plane motion with coordinates r and ϑ. There are two first integrals of this differential equation, given by the two constants of motion

$$l = mr^2\,\dot{\vartheta}^2 \qquad \text{(angular momentum)}$$

with $m = \dfrac{m_1 m_2}{m_1 + m_2}$ and

$$E = \frac{m}{2}(\dot{r}^2 + r^2\dot{\vartheta}^2) \quad \text{(energy)}\ .$$

The explicit solution of (1) reads

$$t = \int_{r_0}^{r} dr' \left\{ \frac{2}{m}(E + \frac{G m_1 m_2}{r'} - \frac{l^2}{2mr'^2} \right\}^{-1/2} \quad , \quad \vartheta = \int_0^t \frac{dt'}{r^2(t')} + \vartheta_0$$

leading to the functions $r(t)$and $\vartheta(t)$ with two arbitrary constants r_0 and ϑ_0 representing the initial values

$$r(t=0) = r_0\ , \quad \vartheta(t=0) = \vartheta_0\ . \tag{2}$$

Hence we obtain the solution in its final form

$$\begin{aligned} r &= f(t; r_0, \vartheta_0; G) \\ \vartheta &= g(t; r_0, \vartheta_0; G) \end{aligned} \tag{3}$$

which describes the real and observable planetary motion.

This simple example illustrates that in the framework of physics the description of the real process $\{r(t), \vartheta(t)\}$ given by (3) is split into three components, the dynamical law (1), the coupling constant G, and the initial values (2). There are no indications in the real planetary motion that a splitting of this kind is meaningful. It is rather induced by our way to describe the real world. The pragmatic reasons for separating laws, initial conditions and constants of nature will be discussed in subsequent sections.

8.1.4 Is the Separation into Laws and Boundary Conditions Necessary?

According to our general understanding of laws of nature these laws should be more general than the particular situation that they describe. In Chap. 5 we emphasised that the laws of nature express invariance properties which are not necessarily present in the real physical process. This concerns in particular the invariance of laws with respect to time reversal. The fundamental laws of nature are – perhaps with one exception[6] – invariant against time reversal, in contrast to the observable real processes that are not time reversal invariant. In the simple example of the preceding Sect. 8.1.3 we find that the law of nature, i.e. the equation of motion (1) is in fact invariant with respect to the time reversal transformation $t \to -t$, but the solution (3), which describes the planetary motion, does not show this invariance. The time reversal symmetry is broken by the boundary condition. If we assume that the values of the spatial coordinates (r, ϑ) are given as *initial values* in the past for $t = 0$, then the future development of the planetary motion can be determined for all values $t > 0$. Similarly, if we assume on the other hand that the values of r and ϑ are given as *final values* in the far distant future, for $t = T$, then the past development of the system is determined for all values $t < T$. Hence, Newton's equation of motion (1) does not define a certain direction of time. We make this distinction by choosing either initial conditions or final conditions.[7]

8.1.5 Classification of Different Kinds of Boundary Conditions – The Propagation of Fields

One might think that the free choice between different kinds of boundary conditions is restricted to the domain of ordinary classical mechanics. This is, however, not the case. In the more fundamental theory of fields which concern classical fields (electrodynamics, gravitation) as well of quantum fields we are confronted with a similar and even more complicated situation. Indeed, for the general Lorentz-invariant field equation we find a large variety of boundary conditions expressing different conceptions of the physical reality.

Let us discuss the relativistic field equation in its most simple form that holds for a scalar field $\psi(\vec{x}, t)$.

$$\Box\,\psi := \left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \Delta\right)\psi = 0 \tag{4}$$

[6] cf. Sect. 7.2.3.4.3, no. (1), Time reversibility.

[7] This point will be further discussed in Sect. 9.1.5.1.

In the presence of sources $f(\vec{x},t)$ we have to consider the inhomogeneous field equation

$$\Box\,\psi(\vec{x},t) = f(\vec{x},t) \tag{5}$$

Equation (5) can formally be solved by means of Green functions according to

$$\psi(\vec{x},t) = \int d^3x'dt'G(\vec{x},t;\vec{x}',t')f(\vec{x},t) \tag{6}$$

where $G(\vec{x},t;\vec{x}',t')$ is the "Green function" that is determined by the differential equation

$$\Box\,G(\vec{x},t;\vec{x}',t') = \delta(\vec{x}-\vec{x}')\delta(t-t') \tag{7}$$

By means of the Fourier transformation

$$G(\vec{x},t;\vec{x}',t') = \frac{1}{(2\pi)^2}\int d^3kd\omega\tilde{G}(\vec{k},\omega)e^{i(\vec{k}(\vec{x}-\vec{x}')-\omega(t-t'))}$$

we find $\tilde{G}(\vec{k},\omega) = -\frac{c^2}{(2\pi)^2}\frac{1}{\omega^2-\omega_0^2}$ with $\omega_0 = ck$ and obtain for the Green function

$$G(\vec{x},t;\vec{x}',t') = -\frac{1}{2}\frac{c^2}{(2\pi)^3}\int d^3ke^{i(\vec{k}(\vec{x}-\vec{x}')}\frac{1}{\pi}\int\limits_{-\infty}^{+\infty}d\omega\frac{e^{-i\omega(t-t')}}{\omega^2-\omega_0^2}) \tag{8}$$

Since the integral

$$I(t) = \frac{1}{\pi}\int\limits_{-\infty}^{+\infty}d\omega\frac{e^{-i\omega t}}{\omega^2-\omega_0^2} \tag{9}$$

has two singularities at $\omega = \pm\omega_0$, which must be circumvented, the integration must be performed in the complex plane. There are five different ways for the calculation of the integral (9) and hence five resulting integrals $I(t)$and five corresponding Green functions which are denoted as

G_p	(principle value Green function)
G_{ret}	(retarded Green function)
G_{av}	(advanced Green function)
G_c	(causal Green function)
G_{ac}	(anticausal Green function).

What does this ambiguity of Green functions mean? Green functions connect the value ψ of the field at a given spacetime point $(\vec{x},t)$ with the values f of the sources at other spacetime points $(\vec{x},t)$. In particular, the choice of a special Green function selects those regions of spacetime that determine the values $\psi(\vec{x},t)$ according to formula (6). As an example we consider the (most important) *retarded* Green function

$$G_{ret}(\vec{x}, t; \vec{x}', t') = 0 \qquad \text{if } t < t'$$

and

$$G_{ret}(\vec{x}, t; \vec{x}', t') = \frac{1}{4\pi|\vec{x} - \vec{x}'|}\delta\left(t - t' - \frac{|\vec{x} - \vec{x}'|}{c}\right) \qquad \text{if } t > t' \ .$$

Accordingly, the *retarded* field $\psi_{ret}(\vec{x}, t)$ is determined by the values of the sources f at spacetime points $(\vec{x}', t') = (\vec{x}', t - \frac{|\vec{x} - \vec{x}'|}{c})$. These points are lying on the backward light cone $L^{-}(\vec{x}, t)$ of the spacetime point $(\vec{x}, t)$ – see Fig. 8.1.

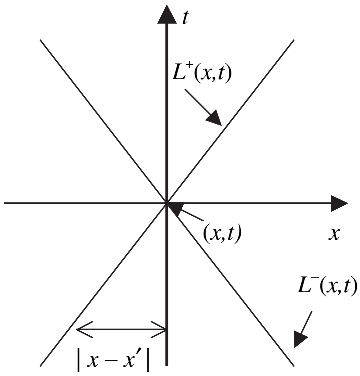


Fig. 8.1. The two light cones of the point (x, t)

The choice of the retarded Green function corresponds to the conception that events in the past determine events in the future. More precisely, the cause should precede the effect, provided the direction of time is established by other physical phenomena (the second law of thermodynamics or the expansion of the universe) The preference for a certain sense of time does not follow from the field equation that is symmetric with respect to time reversal.

If we had chosen the advanced Green function $G_{av}(\vec{x}, t)$ then the sources at points $(\vec{x}', t + \frac{|\vec{x} - \vec{x}'|}{c})$ lying on the forward light cone $L^{+}(\vec{x}, t)$ of the point $(\vec{x}, t)$ would determine the advanced field $\psi_{av}(\vec{x}, t)$, (Fig. 8.1). This choice corresponds to a violation of our concept of causality, since effect precedes cause in time. This is, however, not the right way of speaking. If we were given a time by a process that does not belong to electrodynamics and that flows in the opposite direction, then the advanced Green function would be the right tool for describing the propagation of the field $\psi_{av}(\vec{x}, t)$ – in accordance with our notion of causality.

There is still a third way for describing the physical reality that avoids any preference of a direction of time. It makes use of the Green function calculated

by means of the principle value of the integral $I(t)$ given by (9) that leads to a time-symmetric solution of the field equation given by

$$\psi_p(\vec{x},t) = {}^1\!/_2(\psi_{ret}(\vec{x},t) + \psi_{av}(\vec{x},t)) \; .$$

It has been shown by Wheeler and Feynman[8] that a completely Lorentz-covariant "action at a distance" field theory can be formulated by means of the time-symmetric field $\psi_p(\vec{x},t)$. However, it turns also out that a reformulation of the familiar retarded field $\psi_{ret}(\vec{x},t)$ in terms of the time-symmetric solution requires the hypothetical assumption of many distant "absorbers". We will not go into the details of this rather speculative theory and refer to the literature.[9]

The two remaining Green functions G_c and G_{ac} are discussed in relativistic quantum field theory. They are of less interest for the fundamental point considered in the present section: The Lorentz-invariant field equation is symmetric with respect to time inversion. There are different classes of boundary conditions referring to different directions of time in accordance with the concept of causality. Hence, we have to choose the right class of initial conditions once the sense of time is given by other phenomena. There is no way to read off the direction of time from field theory.

Preliminary Answer to Question 1

Are boundary conditions independent of laws of nature? The two examples, classical particle mechanics and relativistic field theory, show clearly that the separation of the complex real processes into general laws and contingent initial conditions leads indeed to two independent components of our way to describe and to grasp the physical reality. This separation of our experience is man-made and – according to Wigner – perhaps necessary for the human mind to understand the complicated nature.

8.1.6 Are the Laws of Nature Valid also in Other Universes which Differ from Our Universe only with Respect to Initial Conditions?

According to an idea of Popper a law of nature is naturally or physically*necessary* if it holds in all universes that differ from our universe only with respect to initial conditions.[10] An even more general idea is what Weinberg called the symmetry group of nature.[11] The symmetry group of nature is the set of all changes that do not change the laws of nature. Weinberg says that this is the deepest thing that we understand about nature today.[12]

[8] Wheeler, Feynman (1949, DIA).
[9] Wheeler, Feynman (1945, IAR).
[10] Popper (1959, LgF) p. 433. cf. The discussion on the necessity of laws in 9.4.3(3) below.
[11] Weinberg (1987, TFL) p. 72f. Recall also 5.3.2(1) above.
[12] Ibid.

But how can we determine the set of all changes that leave the laws of nature invariant? This would mean to know the line of demarcation between contingent initial conditions and necessary and invariant laws in a most general way (although we know it for large special areas as is clear from 8.1.2 and 8.1.5 above). It would mean to know which constants when changed do not affect the laws and which do; and which initial conditions and boundary conditions would affect the laws when changed and which would not. Are the laws of nature invariant with respect to a slight change of the amount of energy (mass) of the whole universe (which is constant by the law of conservation of energy)? Or could we change the ratio of the electron and proton mass slightly without changing laws. Presumably, this is not the case. Could there be a constellation of the planets (or some star systems) which differs from the one realised now (at a certain time) or in other words: Are different constellations of systems of stars (of our planetary system) at a certain time after the big bang compatible with the laws? In order to answer the question posed in 8.1.6 we shall defend two theses from which it follows that there are possible universes which satisfy all the laws of nature but are different with respect to initial conditions. And this shows in turn that initial conditions must be independent of the laws of nature.

Thesis 1 *The laws of nature (known laws of nature) are valid just in our universe only if the following conditions are satisfied:*

Or in other words: If the laws of nature are valid just in our universe then the following conditions are necessary:

(1) The laws of nature together with at least one initial state are complete with respect to our universe.
(2) All laws of nature are deterministic.
(3) Permutation change (interchange) of elementary particles of the same kind (see 5.3.3(1)) does not change the world (universe).
(4) All states of initial conditions compatible with the laws of nature occur as states (are played through) during the lifetime of our universe.
(5) All fundamental constants are ruled by laws of nature.

Thesis 1 says in other words that all five conditions mentioned above are satisfied if the set of all changes that do not change the laws of nature (the known laws of nature) – is the empty set. Or in other words: All five conditions above are satisfied if the set of all models which are satisfied by the laws of nature is the unit set, i.e. if there is just one model and this is our universe. Einstein's question was more general: Not whether the laws of nature allow more than one world as a model but whether God was free to create another world (even perhaps with other laws).

Thesis 2 *The above five conditions are not (all) satisfied, i.e. the laws of nature (the known laws of nature) are valid also in other universes which differ from ours.*

If thesis 2 is correct, then the initial conditions are independent of the laws of nature. In order to support thesis 2 we have to show that at least one of the above five conditions is not satisfied.

Condition (1): Completeness
As will be said in Sect. 11.1.3 a system L of laws about a certain part of reality, in this case about the whole universe U, is complete, if every truth[13] about U is derivable from L. First of all and more trivially L cannot be complete concerning particular contingent truths about U since L does not contain initial conditions. Thus let us ask whether L plus all initial conditions determining particular states of U in the past ("I_p" for short) is complete.

In Sect. 11.1.3 it will be shown in detail that only for dynamical (and deterministic) laws satisfying conditions D1–D4 (Sect. 7.2.3.2) – in the ideal sense of Laplace – there can be completeness concerning the laws; although also in this case incompleteness concerning initial conditions remains (for example the question about a first state of the system or universe). In many cases like that of dynamical laws which are satisfying condition D4 either not at all or only partially (cf. 9.4) at of statistical laws (cf. 9.4.3.1, S-predictability, and 11.1.3.4) we cannot have completeness.[14] Since we do not assume that the universe is ruled exclusively by dynamical deterministic laws – its expansion, radiation and its thermodynamic processes support this – condition (1) is not fully satisfied.

Condition (2): Deterministic laws
The phenomena of thermodynamics, quantum mechanics, radiation and the new discovered processes of cosmological evolution, of self-organisation and of (certain kinds) of chaos suggest that condition (2) is not satisfied. Of course, this holds under the assumption, which we accept here, that not all randomness comes ultimately from hidden parameters guided by deterministic laws. That means that statistical laws in the realistic interpretation, i.e. describing branches and degrees of freedom in nature, and not just degrees of our ignorance, can be genuine laws of nature (cf. Chaps. 2 and 7). But if this is true then there are degrees of freedom for the development of the universe in the future. And this means that more than one universe are models of these laws.

[13] In order not to run into some logical difficulties (paradoxical situations) we do not permit here a set with the usual logical closure. That means that from a certain specific truth about U (say that proton and neutron have the same intrinsic spin, or that the fine structure constant is about 1/137) not every consequence which is allowed by logic is permitted in this set of truths about the world. Since logic allows a lot of redundant truths to be derived from one true sentence like "p or q" as a logical consequence of p (where q is anything whatsoever). Therefore we restrict the logical consequences by some suitable relevance criterion which has been applied successfully to many different areas like explanation and confirmation theory, verisimilitude (theory of approximation to truth), quantum logic, and still other areas like epistemic and denotic logic. cf. Sect. 3.3.2(b).

[14] For the question of incompleteness of the laws of quantum mechanics cf. 11.1.3.7.

Condition (3): Permutational change
Permutation change, i.e. interchange of elementary particles of the same kind does not change laws but also – according to the usual understanding of a physical system (this system may be the whole universe) – does not change this system (cf. 5.3.3(1)). That means that elementary particles of the same kind are treated as indistinguishable although there are of course more than one particle.

Thus an interchange of two particles of the same kind does not lead to a different world (universe). And thus thesis 2 cannot be supported by permutation change.

Condition (4): Initial conditions
Can we imagine now that all possible initial conditions are or will be realised during the lifetime of the universe? We think that this is not very probable. Even if the choice of initial conditions at the beginning of the universe would have been rather small – a question which is hardly decidable – as soon as we take statistical laws seriously (cf. condition (2)) there will be a great number of states which have not been realised because of the degrees of freedom which allow different states by chance. But even if taking just deterministic laws, with asymmetrical initial conditions asymmetric effects are produced: this special plane of the orbit of a planet (note that it is a plane which follows from the rotationally symmetric laws) is due to initial conditions and the plane could lie in a different angle to the one realised.

Slight changes in the constellations of stars (and planets) seem not to violate laws because such changes occur since planetary motion (and probably this is similar with systems of stars) is to some extent chaotic (cf. 9.4.3.2(3)).

Or take a charge-symmetry violation. The ratio of the decay rate could be slightly different (due to some change in the initial conditions) from the one observed. Take parity: The ratio of the rates of snails having left screw shells to those having right screw shells (or the respective ratio of heart on the left and on the right side) could be different without affecting biological laws. In radioactive decay phenomena parity violation could be more frequently than observed.

Such examples (which could be continued) suggest that it is highly improbable that all possible changes of initial conditions will be realised some time in this universe.[15] That means that if not all initial conditions are played trough during the lifetime of the universe, which is assumed to be finite, then there are also other possible universes satisfying the laws of nature and having some of those initial conditions which are not realised in our universe.

Condition (5): Fundamental constants of nature
This question, whether the fundamental constants of nature are independent of the laws of laws of nature, will be discussed in the subsequent Sect. 8.2.

[15] Recall the discussion of a cosmological question in 7.2.3.4.3(2b).

But since this question is hardly decidable, no answer to it can be used here for a support of thesis 2.

Summarising the discussion of the five conditions of thesis 1 we can say that the conditions (1), (2), and (4) are not fully satisfied. Concerning (4) we have to add the proviso with the universe finite in time. Since some (at least one) of the five conditions are not satisfied, thesis 2 is strongly supported, i.e. the laws of nature are valid also in other possible universes which differ from ours with respect to initial conditions. And this proves that the initial conditions are independent of the laws of nature.

8.1.7 Answer to the Objections

(1) Even if Mach's principle is correct, it does not follow from it that there are no degrees of freedom in the development of the universe. This means that even then condition (4) of the thesis 1 in 8.1.6 can be violated such that not all possible initial conditions are realised. Moreover, some possible microstates of the whole universe will never be realised assuming that the universe has a finite age (cf. 7.2.3.4.3(2b)). Therefore, initial conditions can still be independent of the laws of nature.

(2) Although there is only one universe there are universes possible (not actual) that differ from ours only with respect to initial conditions as has been supported in 8.1.6. Since the possible initial conditions and the possible microstates, which are thought to be realisable are not actually realised in our universe, the laws of nature are valid also in those other possible universes which proves that the initial conditions are not dependent on the laws of nature.

8.2 Are the Constants of Nature Independent of the Laws of Nature?

Terminological remark: Constants in physics and in natural science in general can be divided into two groups. Into the so-called *material constants* and *system constants* on the one hand and into the so-called on the other hand. The first group includes constants like the bulk modulus, the modulus of elasticity, the shear modulus, the Poisson ratio and also constants like the gas constant or the lattice constant of a crystal, etc. The *fundamental constants—f* fundamental constants, on the other hand, include constants like the velocity of light in vacuum (c), Planck's constant ($\hbar$), the gravitational constant (G), the elementary charge (e), the masses of elementary particles like proton and electron (m_P, m_e), Avogadro's constant (N_A), and the Boltzmann constant (k_B). In particular, the fundamental constants include the dimensionless combinations of these constants, the fine structure constant $\alpha = e^2/\hbar c \approx 1/137$ and the ratio of the mass of the proton to that of the electron, m_P/m_e which is approximately 1836.

The above question is concerned with the fundamental constants only. Also the constants of nature discussed subsequently are only the fundamental constants.

8.2.1 Arguments Pro and Contra

8.2.1.1 According to the preceding Sect. 8.1, the initial conditions are independent of the laws of nature. But as Wheeler says, the constants of nature might be more reasonable interpreted as initial conditions:

> "A century and a half ago Laplace dramatized the difference between initial conditions and dynamic laws. The intervening decades have seen new laws uncovered, but not a single discovery about what fixes the initial conditions. The time has come to ask if the constants and the scale of the big numbers belong in the realm of law at all. Are they not more reasonable to be understood as initial conditions?"[16]

Therefore, the constants of nature are independent of the laws of nature provided they can be understood as initial conditions.

8.2.1.2 According to the law of the conservation of energy, the numerical amount of energy of a closed system is constant. This holds also for the whole universe as a closed system. But as it is known from classical mechanics, we can express initial conditions by conserved quantities and vice versa. Therefore such constants are independent of the laws of nature.

8.2.1.3 Measurement units like cm, g, s for length, mass, and time are conventional and thus independent of laws of nature. But many fundamental constants like the Bohr radius, the velocity of light, and the proton mass are measurement units with the help of which all the above traditional units for length, mass, and time and moreover those of the Planck scale can be defined. Therefore such fundamental constants are independent of the laws of nature.

8.2.1.4 No law of nature describes the rate of change of a fundamental constant. But since a change in time of some fundamental constant is possible, the fundamental constants must be independent of the laws of nature.

8.2.1.5 Even though some fundamental constants might not be independent of the laws of nature, a suitable combination of them may still be independent. Thus even if a change of $\hbar$, e, or c would change the laws in which these constants occur, the combination $\alpha = e^2/\hbar c$ could outbalance these changes such that the dimensionless quantity α (fine structure constant) is still independent of the laws of nature. Therefore not all fundamental constants seem to be dependent on the laws of nature.

8.2.1.6 Since the gravitational constant G enters in some of the most basic laws of nature, as those of general relativity, a change of G would change these laws

[16] Wheeler (1973, FRM) p. 241.

of nature such that their time translation invariance would be violated. But according to Dirac's large numbers hypothesis G changes (decreases) with the time development of the universe. Therefore – if Dirac's large numbers hypothesis is correct – some basic laws of nature are not time translation invariant as a consequence of their dependence on constants of nature.

8.2.2 Proposed Answer

> "With the question of the universal constants, you have broached one of the most interesting questions that may be asked at all."[17]

> "In a reasonable theory there are no (dimensionless) numbers whose values are only empirically determinable."[18]

> "Dimensionless constants in the laws of nature, which from the purely logical point of view can just as well have different values, should not exist. To me, with my 'trust in God' this appears to be evident, but there will be few who are of the same opinion."[19]

> "My conclusion is that not only the laws of nature but the constants of nature can be deduced from epistemological considerations, so that we can have a priori knowledge of them."[20]

> "The question of the constancy of such dimensionless numbers is to be settled not by definition but by measurements."[21]

The above quotations are to show that questions about the constants of nature have been a controversial matter. We can add that they are still controversial today. Since our concern are the laws of nature we shall not deal with the constants of nature in general, but we shall comment on three questions which are connected with laws of nature in a more special way. These questions are the following ones: What is the deeper reason for the values of the constants? (8.2.2.1) Are these constants really constant? (8.2.2.2) Are these constants independent of the laws of nature? (8.2.2.3) Honestly speaking, there is no clear answer to the first two questions, there are only conjectures. However, there is an answer to the third question, which will be given below.

[17] Einstein in a letter to Ilse Rosenthal-Schneider of May 11, 1945. cf. Rosenthal-Schneider (1980, RST).

[18] Einstein in a letter to Ilse Rosenthal-Schneider of October 13, 1945. cf. Rosenthal-Schneider (1949, PAE), p. 144.

[19] Einstein in a letter to Ilse Rosenthal-Schneider of March 24, 1950. cf. Rosenthal-Schneider (1980, RST).

[20] Eddington (1939, PPS), p. 58.

[21] Brans, Dicke (1961, MPR).

8.2.2.1 What is the Deeper Reason for the Values of the Constants of Nature?

Before we mention some conjectures concerning this question, we want to emphasise that what is listed under fundamental constants of nature differs to some extent, although the following six are included in most of the serious treatments on constants of nature:

c	velocity of light in vacuum	299 792 458 ms^{-1}
e	unit of electric charge	$1.602\,177\,33 \cdot 10^{-19}$C
m_e	rest mass of electron	$9.109\,3897 \cdot 10^{-31}$ kg
m_p	rest mass of proton	$1.672\,623 \cdot 10^{-27}$ kg
h	Planck's constant (or $\hbar = h/2\pi$)	$6.626\,0755 \cdot 10^{-34}$ Js
G	constant of gravitation	$6.672\,59 \cdot 10^{-11}\ m^3\mathrm{kg}^{-1}\mathrm{s}^{-2}$

Petley[22] adds N_A (Avogadro's constant), k_B (Boltzmann's constant) and μ_0, ε_0 (the magnetic and electric field constants). Dyson[23] adds g (Fermi's constant of weak interaction), H (Hubble's constant) and ρ (the mean density of mass in the universe). Misner, Thorne and Wheeler[24] add the Bohr radius ($a_0 = \hbar^2/m_e e^2$), the reduced Compton wavelength ($\lambda = \hbar/m_e c$) of the electron, the classical electron radius ($r_0 = e^2/m_e c^2$) and the atomic energy unit (e^2/a_B), although they can all be defined with the help of $e, \hbar, m_e$ and c. It should be mentioned that the values of H and ρ have a low degree of reliability. H is uncertain by a factor of about 2, whereas ρ is uncertain by a factor of at least 10^3 because of the invisible mass in the universe. Moreover, in general relativity H is not exactly a constant. Still more difficult is the value of the cosmological constant Λ.[25] Weinberg[26] lists the following three groups:

> "The parameters that appear at the most fundamental level in our present theories of elementary particles are
> (1) the electroweak and strong gauge couplings,
> (2) the masses and self-couplings of the 'Higgs' scalars, and
> (3) the coupling constants for the interaction of the scalars to quarks and leptons."

Concerning the deeper reason for the values of the constants of nature we may distinguish three types of proposals: logical or mathematical ones (1), empirical ones (2), and theoretical and dynamical ones (3).

(1) Logical and mathematical reasons
One version of such a view is represented by Eddington, which is expressed by

[22] Petley (1999, FdK), p. 429. cf. Petley (1988, FCF).
[23] Dyson (1972, FCT), p. 213.
[24] Misner, Thorne, Wheeler (1973, Grav) back cover.
[25] For discussion of the problems see Abbott, L. (1988, RKK) and Weinberg (1989, CCP).
[26] Weinberg (1983, OTP), p. 249.

the quotations above (8.2.2). Eddington was in fact of the rather extreme opinion that the number of protons in the visible universe is deducible from purely "epistemological" or "theoretical" or logical and mathematical premises. This number is also called the "Eddington Number". A more modest view in a similar direction is expressed by Dirac:

> "At present, we do not know why they should have the values they have, but still one feels that there must be some explanation for them and when our science is developed sufficiently, we shall be able to calculate them."[27]

In other views the logical and mathematical reasons are complemented by additional aesthetic reasons like that of simplicity. Such a view is present in the following famous quotation of Einstein: "What I am really interested in is whether God could have made the world in a different way; that is, whether the necessity of logical simplicity leaves any freedom at all."[28]

(2) Empirical reasons
The empirical reasons for the constants of nature may be divided again into two types of proposals: non-teleological ones and teleological ones. An example of a non-teleological view is some version of what is called *Mach's principle*: The whole matter in the universe and its distribution determines the inertial field (the guiding field) of all motions and consequently all the laws of motion and presumably also some constants of nature.[29]

Examples of a teleological view are all versions of the so-called *anthropic principle*. The following principle is what Barrow and Tipler call the "weak anthropic principle":

> "The observed values of all physical and cosmological quantities are not equally probable but they take on values restricted by the requirement that there exist sites where carbon-based life can evolve and by the requirement that the Universe be old old enough for it to have already done so."[30]

(3) Theoretical and dynamical reasons
Within the framework of a universally valid quantum field theory – a "final theory" as it was called by Weinberg – it should be possible in principle to calculating not only some masses of elementary particles but also various coupling constants between these particles. Although a "final theory" of this kind

---

[27] Dirac (1973, FCD), p. 45.
[28] Hawking, Israel (1987, ECV), p. 128.
[29] The constants of nature are, according to such an interpretation, properties of matter, whereas the laws of nature, if they are correct, describe all motions. For an elaboration of Mach's principle especially as it is manifest already in Aristotle, Copernicus and Kepler and for Mach's critique of Newton, see Barbour (1989, ARM). cf. also Sects. 6.2.1(6) and 6.2.2(4).
[30] Barrow, Tipler (1986, ACP), p. 16 and Barrow (2002, CNt), Chap. 8.

is not yet in sight several attempts were made to calculate a mass spectrum for elementary particles and coupling constants. For example, Heisenberg tried to determine the value of the fine structure constant by means of his non-linear spinor theory of elementary particles.[31]

8.2.2.2 Are the Constants of Nature Really Constant?

On a more general point of view we have to say first that anybody, who would answer the above question with "No", has to provide another measuring rod that is more stable than the magnitude that is to be investigated. In other words, a change of some fundamental constants is only meaningful if at least some other constants remain unchanged. This leads to the following aspects of the above question: Stability of the measuring rods and accuracy of measurement (1), reduction of some of the constants to others (2), what are the consequences if some constants would vary? (3).

(1) Stability of the measuring rods and accuracy of measurement
Maxwell had already made an important observation on this aspect:

> "Yet, after all, the dimensions of our earth and its time of rotation, though, relatively to our present means of comparison, very permanent, are not so by any physical necessity ...
>
> But a molecule,[32] say of hydrogen, if either its mass or its time of vibration were to be altered in the least, would no longer be a molecule of hydrogen.
>
> If, then, we wish to obtain standards of length, time and mass which shall be absolutely permanent, we must seek them not in the dimensions, or the motion, or the mass of our planet, but in the wavelength, the period of vibration, and the mass of these imperishable and unalterable and perfectly similar molecules."

Maxwell's idea is realised to a great extend by interpreting e, m_e, m_p, $\alpha = e^2/\hbar c$, and m_p/m_e as very stable and permanent. The values of α and m_p/m_e have a constancy of at most 1% during the lifetime of the universe so far (if the other usual underlying calculations are correct).[33] Within the last decades the accuracy of measurement for constants has been raised for one decimal place within every period of 15 years. This was possible in many cases by new experimental techniques and by extensive use of computers. Moreover it should be mentioned that the magnitudes m_p/m_e, e, $\hbar$ and the *Rydberg* constant, furthermore some auxiliary magnitudes like the mol-mass of the proton are known today to a high degree of accuracy (10^8 or more). Presently, there

[31] Heisenberg (1967, ETE), p. 120 equation 8(32).

[32] What Maxwell has in mind is in today's terminology an atom. This passage is from Maxwell's Presidential Address to the British Association for the Advancement of Science of 1870, quoted in Petley (1985, FPC), p. 15.

[33] cf. Petley (1999, FdK), p. 430. cf. Shlyakhter's conjecture below.

is still a special difficulty with G because it does not appear in a combination of other fundamental constants, which are measurable more accurately.

(2) Reduction of some of the constants to others
According to Planck some constants are more basic than others such that not everything is conventional in this respect:

> "In contrast with this it might be of interest to note that, with the aid of the two constants h and k which appear in the universal law of radiation, we have the means of establishing units of mass, time and temperature, which are independent of special bodies or substances, which necessarily retain their significance for all times and or all environments, terrestrial and human or otherwise, and which may, therefore, be described as 'natural units'."[34]

Today Planck's units are defined with the help of the constants $\hbar$, c and G as follows:

$(\hbar c/G)^{1/2}$	Planck mass	$2.177 \cdot 10^{-8}\,\mathrm{kg}$
$(G\hbar/c^3)^{1/2}$	Planck length	$1.616 \cdot 10^{-35}\,\mathrm{m}$
$(G\hbar/c^5)^{1/2}$	Planck time	$5.391 \cdot 10^{-44}\,\mathrm{s}$

A further example of a reduction is motivated by the fact that the accuracy of the measurement of the velocity of light c has become much greater such that the unit of length (one meter) is defined now with the help of c.[35]

(3) What are the consequences if some constants would vary?
The constants which are usually regarded as the defining fundamental properties of the physical universe are those listed by Dyson (1972, FCT) p. 213, i.e.: c, $\hbar$, e, m_p (or m_p/m_e), g (Fermi's constant), G, H (Hubble's constant) and ρ (mass density of the universe). From these quantities one can construct five dimensionless constants as follows:

$$\begin{aligned}
\alpha &= e^2/\hbar c &&= 7.297\,353\,08 \cdot 10^{-3} (\approx 1/137)\\
\beta &= g m_p^2 c/\hbar^3 &&= 9.0 \cdot 10^{-6}\\
\gamma &= G m_p^2/\hbar c &&\approx 5 \cdot 10^{-39}\\
\delta &= H\hbar/m_p c^2 &&\approx 10^{-42}\\
\varepsilon &= G\rho/H^2 &&\approx 2 \cdot 10^{-3}
\end{aligned}$$

Dirac's original question of (1937, CCs) and (1938, NBC) was devoted to the problem, which ones of these dimensionless constants vary with the evolution of the universe. There are five different hypotheses, which give an answer to that question.[36]

---

[34] Planck (1913, THR), p. 175.

[35] Until 1960 the "metre" was defined by the wave length of a particular spectral line of the krypton atom, since 1983 the "metre" is defined by the velocity of light c and the time unit (second) based on atomic clocks.

[36] For a detailed description see Dyson (1972, FCT).

Before we consider these hypotheses we can make some general remarks about α and m_p/m_e: First, both numbers are essential for the structure of atoms and molecules; a variation of them would change this structure correspondingly. According to Shlyakhter's conjecture the rate of change of α compared to the value of α is $< 5 \cdot 10^{-18}$ per year.[37] As has been said already, m_p/m_e and α can be measured with a very high degree of accuracy. From this degree of stability of α it follows immediately that those hypotheses according to which α varies with time (with the evolution of the universe) to a higher degree must be excluded.

Dirac's Hypothesis

In 1937 Dirac proposed a *numerical principle* according to which very large or very small numbers should not occur in the basic laws of physics. Since γ and δ are extremely small magnitudes, Dirac's hypothesis assumes α, β, ε to be constant but γ and δ to vary with time. The two main consequences of this hypothesis are, that gravitational forces (measured via γ) decrease with time and since ε is constant and ρ varies with time as t^{-1} (and not as t^{-3}) spacetime has zero curvature.

In 1972 Dirac formulated his large numbers hypothesis,which has similar consequences. This hypothesis states that very large numbers have to be compared in some way.

> "It involves the fundamental assumption that these enormous numbers are connected with each other. The assumption should be extended to assert that, whenever we have an enormous number turning up in nature, it should be connected to the epoch and should, therefore, vary as t varies. I will call this the Large Numbers Hypothesis."[38]

Dirac found another large dimensionless constant which is the ratio of the electric force e^2/r^2 between electron and proton and the gravitational force $G \cdot m_p \cdot m_e/r^2$ between electron and proton, the quantity $e^2/G \cdot m_p \cdot m_e$ which is dimensionless and of the order of about 10^{40}. He compared this number with the age $T$of the universe in terms of atomic units, for example expressed in time units $t_e = r_e/c$, where t_e is the time the light needs for the distance of the diameter of the electron. The age of the universe (as known today) in time units t_e is also about 10^{40}. Thus he proposed the equation

$$e^2/G \cdot m_e \cdot m_p = T/t_e \tag{*}$$

as a fundamental equation expressing his so-called "large numbers hypothesis".

This hypothesis and more specifically the above mentioned equation (*) that connects the gravitational constant with the age of the universe has severe consequences:

---

[37] cf. Dyson (1978, VCs), p. 164, and Damour, Dyson (1996, OBT).

[38] Dirac (1973, FCD), p. 46.

(1) If this equation is true, then the laws of nature are not time translation invariant. According to Dirac $\dot{G} \neq 0$, and G should decrease with time. A further consequence would be that the law of the conservation of energy would not hold any more.
(2) As Dirac points out, the big bang theory, when developed in accordance with the large numbers hypothesis, implies continuous creation of matter, which violates the law of conservation of energy.

A consequence of (1), i.e. $\dot{G} \neq 0$, would be that the moon should depart from the earth in the course of time and that some of the constants in the above equation – or respectively in the following transformation of it – would not be really constants: $T = e^4/m_e^2 \cdot m_p \cdot G \cdot c^3$. The most exact measurements concerning a departure of the moon did not show a significant effect or, more accurately, are compatible with both $\dot{G} = 0$ and the expected deviation.[39]

Summing up we may say that – in the light of the data available today – these, and the above stated consequences, do not make Dirac's hypothesis very probable.

Brans–Dicke Hypothesis

The *Brans–Dicke* hypothesis is very similar to that of Dirac.[40] It also assumes α, β and ε constant but γ and δ to vary with time. But the difference is that in Dirac's hypothesis γ varies with t^{-1}, whereas in the Brans–Dicke hypothesis it varies with t^{-r}, where r is a small number of the order of 0.05. Therefore the Brans–Dicke hypothesis implies a much weaker time variation of the gravitational forces than Dirac's hypothesis.

Conventional Hypothesis

The conventional view assumes α, β, γ to be constant and δ and ε to vary with time as t^{-1}. This view is based on the idea that the five dimensionless magnitudes are divided into the "laboratory constants" α, β and γ and the "cosmological constants" δ and ε. The latter are assumed to vary with time and the universe is described by an open cosmology in which the distance between two separated galaxies varies linearly with t.

There are two further hypothesis, that of Gamov (β, γ, ε constant, α, δ varying with time) and that of Teller, Landau and DeWitt (β, ε constant, α, γ, δ varying with time, α logarithmically) which can be taken to be already refuted by experience: Gamov's hypothesis is excluded by direct astronomical measurement of α, while the hypothesis of Teller, Landau and DeWitt is excluded by geophysical isotope abundances.[41]

---

[39] cf. Dyson (1972, FCT), p. 227; Misner, Thorne, Wheeler (1973, Grav) §40,8, p. 1121ff; Irvine (1983, CLP), p. 429; Rees (1983, LNR, 2002, NCT); Damour et al. (1988, LVG); Genz, Decker (1991, SSB), p. 94; Damour, Dyson (1996, OBT).

[40] For the Brans–Dicke hypothesis there is also a theoretical justification. cf. Brans, Dicke (1961, MPR).

[41] cf. Dyson (1972, FCT), p. 235.

8.2.2.3 Are the Constants of Nature Independent of the Laws of Nature?

We say that the constants of nature are independent of the laws of nature if either a change of these constants does not change the laws or if a change of the laws does nor change the constants. A rough answer to the question 8.2.2.3 is as follows: Those constants of nature which enter (are constituents of) laws of nature are not independent of laws of nature. Because a change of these constants in time provides a change of the laws. For example a decrease of G in time implies that the respective laws are no more time translation invariant. Now since $\hbar$, e, m_e, c and G enter laws of nature, these constants are not independent of laws of nature. – In addition, those constants which can be deduced from laws of nature in the sense of Heisenberg, say, are also not independent of the laws of nature.

From a more detailed point of view we have to distinguish constants which directly enter laws (1) from those which indirectly enter laws (2). The latter may be divided into those (2a) which result from combinations of (1); further into those (2b) which are scale constraints and into those (2c) which are first initial conditions at the origin of the universe and which (might) determine the constants of nature.

(1) The constants of nature which directly enter laws of nature are threefold: First, there are those which enter the laws of non-relativistic quantum mechanics: $\hbar$, m_e and e. A change of any one of them would cause a change in the laws; for example a change of $\hbar$ would provide a change of Schrödinger's equation.

We have to observe however that "change" can mean two different things: a change of $\hbar$ can mean that $\hbar$ has another value; and in this case the original Schrödinger equation would not be valid anymore. A change of $\hbar$ could also means that $\hbar$ changes with time; and in this case Schrödinger's equation would be not enough detailed, it would have to take this change into account. With respect to both cases of "change" however these constants of nature are not independent of the respective laws of nature.

Secondly, there is the constant that enters Maxwell's theory and the special relativity theory, the velocity of light c. Both types of changes mentioned above would affect the respective laws. For example a change of c would affect Maxwell's equations and the Lorentz transformations.

Thirdly, there are those constants which enter the field equations of general relativity theory: c and G. It was already mentioned that a change of G with time implies that the respective laws of nature are not anymore time translation invariant.[42]

(2a) Some of the most important derived constants which can be defined with the help of the fundamental constants listed in 8.2.2.1 are α (the fine structure constant) and the ratio of proton and electron mass $m_p/m_e \approx 1836$.

[42] cf. note 34 above and the comments to Dirac's Large Numbers Hypothesis.

As mentioned above (8.2.2.2(3)), they can be measured with great accuracy and are constant at least to a very high degree. Further important derived constants are the Rydberg constant, which also plays an important role in spectroscopy, and the Bohr radius a_0 . It is obvious that these constants are not independent of the laws of nature. Any change of them would affect the basic laws of nature.

(2b) Constants which are scale constraints are Planck's units (cf. 8.2.2.2(2) above). The Planck units are scale constraints in the sense that at these values a breakdown of spacetime structure is to be expected: The laws of general relativity theory would not hold anymore such that this theory has to be replaced by a theory of quantum gravity yet to be found. From this it is clear that a change of Planck's units would affect laws of nature. In this sense also the constants $\hbar$ and c are scale constraints. Planck's constant $\hbar$ defines the borderline between quantum mechanics and classical mechanics and c defines the border line between Special Relativity and classical mechanics. Hence a change of $\hbar$ and c would shift the interface between classical mechanics on the one hand and quantum theory and special relativity on the other hand. (For consequences of changing $\hbar$ and c see the nice story by Gamow, G. (1965 MTP))

(2c) Assuming a hot big bang cosmology there is the question whether some initial conditions at the beginning of the universe might determine fundamental constants of nature.[43] Thus for example the ratio Ω_0 of the potential energy of the universe and its kinetic energy of expansion, which equals the ratio ρ_0/ρ_c of the matter density ρ_0 of the universe, and its critical density ρ_c (i.e. the largest density the universe can possess and still expand) determines G and H (Hubble's constant) via the equation $\rho_0/\rho_c = 8\pi G/3H^2$, which holds in one of the Friedmann models. According to today's calculations the universe is very close to the critical state with $\Omega_0 = 1$. This implies a special expansion rate at Planck time which again determines H and G.

Another example is this:[44] One starts from the experimental fact of the cosmic background radiation (Penzias and Wilson, 1965) with its present temperature of about 3 K. Its thermal spectrum[45] corresponds to a certain black body radiation density which leads to the ratio of baryon to photon in the universe. This baryon photon ratio is another important cosmological parameter which determines Ω_0 and the cosmic entropy per baryon.

To sum up we can say that certain initial conditions at the beginning of the universe might determine certain constants of nature. Now, although initial conditions in the usual sense are independent of laws of nature (recall Sect. 8.1), the question is more difficult concerning those initial conditions at the beginning of the universe. Since if they are determining constants of nature we cannot say that they are independent of the laws of nature.

---

[43] cf. Barrow, Tipler (1986, ACP), ch. 6, and Rees (1983, LNR), p. 313ff.
[44] cf. ibid. p. 380.
[45] Calculated by Woody, Richards (1979, CBR).

From a more critical point of view, however, we have to add that nobody knows exactly what is determined first at the beginning of the universe (around the area of Planck time): the laws, the constants, or the initial conditions. Therefore, a strict answer to the question whether some constants of nature are dependent on initial conditions at the beginning of the universe is not possible.

8.2.3 Answer to the Objections

8.2.3.1 (to 8.2.1.1) The answer to this question is clear from 8.2.2.3: Constants of nature are not independent of laws of nature. According to 8.2.2.3(2c) one has even to add a proviso concerning those initial conditions at the beginning of the universe which might determine constants of nature.

8.2.3.2 (to 8.2.1.2) Concerning this objection we have to distinguish two questions: First (1), whether a magnitude which is conserved according to a conservation law is rather an initial condition than a constant. Second (2), whether a conserved magnitude which is a parameter of the whole universe may still determine constants of nature (constants of the universe).

(1) In many cases a magnitude which is conserved according to a conservation law turns out to be rather an initial condition than a constant. This can be seen from the following example:
   Consider in classical mechanics the so-called harmonic oscillator $m\ddot{x} + m\omega^2 x = 0$. Then there are two possibilities:
   (a) We are given the initial values $x(0) = x_0$ and $\dot{x}(0) = v_0$.
   (b) We are given the energy $E = \frac{m}{2}\omega^2(x_0^2 + \frac{v_0^2}{\omega^2})$ together with the phase φ of the oscillation. Since we have $x_0/v_0 = (1/\omega)tg\varphi$, we can calculate the initial values with the help of E and φ. Therefore the values of such conserved magnitudes behave like initial conditions. And as initial conditions they are independent of the laws of nature.
(2) On the other hand a magnitude like the total energy of the universe or its total mass may determine fundamental constants. Thus if the total mass M of the visible universe is given by

$$M = (4\pi/3)\rho(ct_u)^3 \approx c^3 t_u/G ,$$

one can see – keeping t_u and G the same – that c could be greater if M were greater.[46] A similar consideration of inverse dependence can be made with G.

[46] A related idea has been proposed by Meessen (2000, STQ). According to his theory of spacetime quantisation $c = 2aE_u/h$, (where a is the smallest measurable distance but not necessarily identical with the Planck length). Here c depends on the amount of energy E_u of the universe and may be different if this amount is different – although there are serious objections against such a view.

> On a more critical point of view one has to keep in mind that calculations of the total energy or mass of the (observable) universe are some kind of extrapolation from present cosmological data. Strictly speaking we do not have a balance to weigh the universe if it is closed: "Around a closed universe there is no place to put a test object or gyroscope into Keplerian orbit to determine either any so-called 'total mass' or 'rest frame' or '4-momentum' or 'angular momentum' of the system."[47]

8.2.3.3 (to 8.2.1.3) Among the fundamental constants of nature we have to distinguish two types: First those which serve as a basis for defining measurement units and secondly those which cannot be used that way.

To the first group belong e, m_e, m_p, $\hbar$, ε_0, c and G. For example a mass unit can be defined just as m_e or m_p and a length, time and amperage unit with the help of $\hbar$, m_e, e and ε_0. The Planck units can be defined with the help of $\hbar$, c and G. In this sense these fundamental constants can be viewed as natural measurement units with the help of which other conventional measurement units are definable.

To the second group belong α, m_p/m_e, β, γ (recall 8.2.2.2(3)). They cannot be used to define conventional measurement units.

Now concerning the question of dependence or independence of the constants from laws of nature it is easily seen that the fact that these constants can define conventional measurement units does not make them conventional in the same sense; and moreover this fact is not sufficient to make them independent from laws of nature. As is clear from 8.2.2.1 a change of e, $\hbar$, m_e, m_p, c, G would affect laws of nature. This holds also for those constants which indirectly enter laws, like α or m_p/m_e. Therefore the conclusion of the argument 8.2.1.3 is not correct.

8.2.3.4 (to 8.2.1.4) To this argument we have to say two things: First, that the first premise is only correct under a condition and second that these two meanings of "independence" are confused in the argument.

As to the first, the first premise "No law of nature describes the rate of change of a fundamental constant" is correct. But it is correct only under the condition that the fundamental constants are really constant. If some change in time of some of the fundamental constants could be established experimentally, then a law could be formulated which describes the rate of change.

As to the second, the argument does not distinguish two types of "independence" or "dependence":

(a) The independence or dependence of a constant that enters a law, from this law.

[47] Misner, Thorne, Wheeler (1973, Grav), p. 458.

(b) The independence or dependence of a constant from a law which describes the change of this constant, whereas (a) applies always (b) applies only if there is a change of a constant and a respective law to describe it. Now the argument 8.2.1.4 uses (b) in its first premise and (a) in its conclusion. Therefore the conclusion of this argument is not proved.

8.2.3.5 (to 8.2.1.5) As has been said in 8.2.2.3(2a) constants like α or m_p/m_e enter indirectly into the laws of nature. Now there is a difference between such constants as α, β, γ or the Rydberg constant on the one hand and m_p/m_e on the other hand. The first could keep their values although the defining constants change in such a related way as to outbalance a change of the defined constants; thus α could keep its value although e, $\hbar$ and c would change respectively. But from this it does not follow that α is independent of laws of nature. Since a change of α would imply a change of at least one of the defining constants, this would effect those laws of nature where this defining constant (in the case of α: e, $\hbar$ or c) would enter. Moreover, if α were derivable theoretically – in the sense of Heisenberg mentioned above – the fine structure constant would depend on those laws that allow it to be determined. Concerning m_p/m_e, it is obvious that a change would affect laws of nature, since laws of nature are in general not scale invariant (recall Sect. 5.4.1(3)).

8.2.3.6 (to 8.2.1.6) This argument is in agreement with what has been said in Sect. 8.2.2.2 to Dirac's hypothesis.

9

Causality and Predictability

9.1 Do all Laws of Nature Represent a Causal Relation?

Arguments Contra

(a) In dynamical laws, the differential equation can be interpreted as representing a causal relation where the description of the initial state represents the cause, and the description of the final (predicted) state represents the effect. With statistical laws, however, no such interpretation is possible. But, as is clear from Chaps. 2 and 7, statistical laws are genuine laws of nature. Therefore, not all laws of nature represent a causal relation.

(b) Let us assume that every causal relation is asymmetric, i.e. if $C(x, y)$ is the causal relation between the events (states) x and y, then

$$\forall_x \forall_y (C(x, y) \rightarrow \neg C(y, x)) \; ;$$

or, in words: if x causes y, then not: y causes x. But the law of gravitation expresses a symmetrical influence of mutual attraction between two bodies or states. Therefore, if causal relations are assumed to be asymmetric, then not all laws of nature represent a causal relation.

Arguments Pro

Every law that describes a definite dependency of a later state on an earlier state can be interpreted as describing a causal relation. If every law of nature describes a definite dependency of a later state on an earlier state, then every law of nature describes a causal relation.

General Aspects of the Causal Relation Discussed in History

9.1.1 Aristotle

In his metaphysics,[1] Aristotle first gives a definition of cause, then distinguishes four types of causes, and finally states properties of the causal relation.

[1] Aristotle (Met), Book V, Chap. 2.

To obtain a definition of cause, Aristotle first discusses the genus of cause which is "principle" (arché, according to him). After distinguishing six types of arché, he says that every cause is an arché of a certain sort. More accurately, the definition is this:

> Definition: Every cause is an arché, i.e. something which is the first from which either being or coming into being or knowledge starts.
>
> Four types of causes: He then distinguishes four types of causes to which special questions are related:
>
> C_m causa materialis (material cause). Questions: Out of what? From what? Interpretation (in the light of our knowledge today): The chemical element with its isotopes, the material constants such as the elastic modulus of a certain metal, the diameter of a proton (and similar numerical values which are scale invariant) are parts of C_m.
>
> Cf_o causa formalis (formal cause). Question: What is it? Interpretation: The chemical element with its isotopes, the atomic or the crystal structure, the particular species (of a plant or animal), the DNA of a species, the kind of phenomenon (for example: an eclipse, a DNA recombination, etc.)
>
> C_e causa efficiens (efficient cause). Questions: Which is the starting point? Who is the producer? Who is the mover? Interpretation: The parents (with respect to the children), the billiard player, the car driver, the earlier state of the system with respect to the later state where both are connected by a dynamical law.
>
> Cf_i causa finalis (final cause). Questions: Why? For what? What for? Interpretation concerning the question "why?": The premises (with respect to the conclusion), the law (with respect to the explained phenomena). For Aristotle, every scientific explanation is an answer to a "why" question. Interpretation concerning the questions "for what?" and "what for?": The goal, purpose, aim, motive, intention, ... etc. This second interpretation is the usual one which is described as teleological, but one should not forget the first one, which is the important one for scientific explanation.
>
> Properties of the causal relation: The causal relation is always asymmetric:
>
> $\forall_{a,b}(C(a,b) \rightarrow \neg C(b,a))$. But, if the relation is different, then it may hold for some a, b both: $C_e(a,b)$ and $Cf_i(b,a)$. For instance, the master-builder is the causa efficiens of the house, and the house is the causa finalis of the master-builder. The asymmetry of the causal relation implies its irreflexivity (i.e. a cannot be a cause of a). According to Aristotle, the causal relation is sometimes counterfactual in the sense that: For some a, b: If the obtaining of a is C_e for the obtaining of b, then the not-obtaining of a is C_e for the not-obtaining of b. Concerning time, he holds that cause and effect are non-simultaneous if the cause is in potency

(with respect to the effect); however, cause and effect are simultaneous if the cause is in actuality (with respect to the effect).

9.1.2 Thomas Aquinas

One important place to find properties of the causal relation and its application are his proofs for the existence of God in his *Summa Theologica* (part I, qu.2, art.3). The second of the five proofs ("five ways") is explicitly concerned with the order in the efficient causes. There are three axioms explicitly in the text which describe the causal relation of C_e:

A1: Something causes something else.
A2: Nothing causes itself (irreflexivity).
A3: If there is no first cause for a thing, then there is also no intermediate and no final cause for that thing.

In addition, the text defines a first cause (of other things) by stating at least the following properties: It does not cause itself (irreflexivity), it is not caused by anything else, and it is the first in a causal order. Thomas Aquinas concludes from this: It is necessary to assume one first cause. The exact formulation in symbolic language of these axioms and the definition is as follows:

$$\begin{aligned} &\mathrm{A1}: && \exists_x \exists_y C(x,y) && \text{(`}C(x,y)\text{' for `}x\text{ causes }y\text{')}\\ &\mathrm{A2}: && \forall_x \neg C(x,x) &&\\ &\mathrm{A3}: && \forall_y \left[\neg\exists_z FC(z,y) \rightarrow \neg\exists_x C(x,y)\right] &&\\ &\mathit{Def.}: && FC(x,y) \leftrightarrow [x \neq y \wedge \neg C(x,x) \wedge \forall_u \neg C(u,x) &&\\ & && \wedge \exists_{w_1} \ldots \exists_{w_n} (C(x,w_1) \wedge \ldots \wedge C(w_n,y))] && \end{aligned}$$

$$\text{(`}FC(x,y)\text{' for `}x\text{ is the first cause with respect to }y\text{')}$$

In his reconstruction of this "second way", Essler[2] interprets Thomas as intending to prove: There is one thing which is uncaused and causes everything. In order to prove the second part ("...which causes everything"), Essler assumes a fourth axiom, saying that the causal relation is partially connected: If neither x causes y nor y causes x, then there is a z such that z causes both x and y:

$$\mathrm{A4}: \quad \forall_x \forall_y \left[(x \neq y \wedge \neg C(x,y) \wedge \neg C(y,x)) \rightarrow \exists_z (C(z,x) \wedge C(z,y))\right]$$

From A1–A4 plus a simpler but stronger version of the definition of the first cause, Essler shows that one can derive: There is one first cause which is uncaused, which does not cause itself and which causes everything (other than itself). He also believes that Thomas assumes axiom 4, and that the causal relation is transitive, although transitivity is not used in the argument of Aquinas. We think that this interpretation is too strong in the following sense:

[2] Essler (1969, Elg).

(1) There is no claim in the text that the first cause (God) causes everything. To be the first member in a causal order does not mean that the first member causes every member in the order. If parents produce children, and these children produce their children (or commit a crime), we do not say that the parents produced the grandchildren (or committed the crime). That means that the causal relation, in the wide sense, is not generally transitive. However: If the parents do not exist, the grandchildren and their crime do not exist either. This is (when generalised) Thomas' axiom 3, and it tells that the first cause is a necessary condition (but not a sufficient one) for all other members, although it is also a sufficient one at least for the second cause (member). Thomas says explicitly that God neither wants nor causes free immoral actions (otherwise he would be inconsistent having given the ten commandments). Since he is neither an allwilling nor an allcausing God, he does not apply his will or power to every state of affairs.
(2) If God as the first cause does not cause everything, then there is no need for axiom A4.
(3) In the text we do not find asymmetry either. From asymmetry, irreflexivity follows, but not the other way round. Aquinas seems only to require irreflexivity (except for the relation of the first cause to the second or the "higher secondary" causes, where asymmetry must hold additionally). This is interesting, since the fundamental laws of Classical Mechanics and Relativity Theory are time symmetric such that they do not designate a causal order in one direction only. To give an example from another area: The neuronal connections in the brain are reciprocal and do not show an asymmetry either (cf. however 9.2.1(2)).

Summarising, we may say that Thomas Aquinas has a very modest (and wide) concept of the causal relation: It is irreflexive, it holds (at least) between things (or events) of this world (axiom A1), and the net (branch or chain) of causal relations has to have a first element.

9.1.3 Leibniz

The causal principle of Leibniz is his principle of sufficient reason (PSR). As has been shown elsewhere,[3] this principle embodies a claim of the completeness of certain scientific systems which are built up *more geometrico* or of their axioms, respectively.

PSR Nothing happens without sufficient reason or every truth has its proof ... from the axioms and definitions.[4]

To see that PSR is a completeness claim, we have to recall the definition of (semantical) completeness: A set S of axioms (laws) is complete with respect

[3] Weingartner (1983, IMS).
[4] Leibniz (GPh), Vol. 2, p. 62.

to an area of events (effects, phenomena) E, if all true sentences E′ which are descriptions of E are derivable from S.

Leibniz thought that logic, mathematics and metaphysics, built up as axiomatic systems, are complete in the sense above, and men can, in principle, find out the respective axioms. With respect to physics, jurisprudence and ethics, however, he thought that there exist also axiomatic systems for these areas which are complete. But, since some of their theorems are infinitely analytic (i.e. can be traced back to simple axioms only within infinitely many steps), these axiom systems cannot be known (and their completeness cannot be proved) by men. PSR, however, is not only a completeness claim about the respective axiom systems or sufficient premises. It is also one about the completeness of the sufficient reasons or sufficient causes for an area of events: A set C of reasons or causes is complete with respect to an area of events (effects, phenomena) E, if all events e of E can be scientifically explained or are caused by C. In the light of our knowledge today, the question is: Are the reasons or causes complete with respect to the known physical phenomena? Taken generally, the answer is, of course: No. More specifically, there is some suitable completeness with respect to some restricted areas, but there is no completeness with respect to others (chaotic phenomena, for instance).[5]

9.1.4 Newton: Causes Interpreted as Forces

According to Newton's Principia, the causes which produce or change ("true") movements are forces:

> "The causes by which true and relative motions are distinguished, one from the other, are the forces impressed upon bodies to generate motion. True motion is neither generated nor altered, but by some force impressed upon the body moved; but relative motion may be generated or altered without any force impressed upon the body."[6]

Some important consequences for the causal relation, as it is understood by Newton (according to this passage), are as follows:

1. Causal change takes place on the background of noncausal inertial movement.
2. There is causeless change of position. Of the two components of the movement of a planet on its orbit, one, the tangential, is non-causal, and the other, the centripetal, is causal.
3. Cause (force) and effect (acceleration) are simultaneous.
4. Causeless or non-causal movement and change does not mean indeterminism: Inertial movement may be determined, at least in the sense of being predictable without being a matter of chance, but is causeless, according to Newton.

[5] For the question of completeness of the laws of nature see Sec. 11.1.

[6] Newtond (Princ), Book I, Definitions, Scholium.

9.1.5 Newton, Lagrange, Laplace, Hamilton, Maxwell

What Newton began by using comprehensive laws for the explanation of physical phenomena was further developed by Lagrange, Laplace, Hamilton, and Maxwell. It was understood that, since these laws describe the time development of a physical system from one of its earlier states to its later state, this later state could be interpreted as the effect of a cause which is represented by the earlier state. And, the important relation between the two states is the causal relation or connection that is represented by the dynamical law in the form of a differential equation. While Newton's idea to interpret causes as forces did not become influential at all – probably also because of the strange consequences mentioned above – the idea to interpret the causal relation with the help of the dynamical laws of Classical Mechanics began a triumphant march still current these days: The usual understanding a physicist or natural scientist has of causality is a relation between states of a system where this relation is described by a dynamical law. However, this does not mean that this usual understanding is accurate enough to characterise dynamical laws by the conditions D1–D4 (or by D1, D2 and D4) of Sect. 7.2.3.2. Although D1 will be assumed explicitly, there is still ignorance of D4 which is, however, crucial, as has been elaborated in 7.2.3.2 above.

9.1.6 Hume

David Hume raised the question why the concept of *causality* is usually considered as intimately related to the idea of necessity. Where does this connection come from and how can it be justified? Hume did not question that there is a necessary connection between cause and effect. He writes

> "According to my definition, necessity makes an essential part of causation"[7]

and he had no doubts in the existence of causality at all,

> "but allow me to tell you, that I never asserted so absurd Propositions as that any thing might arise without a cause".[8]

Hume has no doubts in the reality of the external world and in the causality which is expressed by the laws of nature.[9] His scepticism is focused on the human capability to recognise the strict causality relation. This problem has two facets. First, a causality proposition C(A, B) states, according to Hume, that in all sequences of events A and B these events are connected necessarily.

[7] Hume, D. (1939, THN), Book II, Part III, Sect. I.
[8] cf. K. Smith (1966, PDH) p. 431 (letter by Hume, February 1754).
[9] That Hume adheres to the necessitarian standpoint and to a realistic position is elaborated in detail by Bonk (1998, KIA). cf. also K. Smith (1966, PDH).

However, this corresponds to an infinite (quantifier) proposition, which cannot be justified by a finite sequence of propositions. Second, Hume assumes that a causality relation tells us something about nature, about the external reality. And here we are confronted with the problem, how this causality in the external reality can be justified by induction.

The question what causes us to believe that causality is intimately connected with necessity is answered at first by a recourse to human customs and habits. This way of reasoning expresses Hume's scepticism against the arguments of the rationalists as well as of the empiricists. However, Hume's final answer is concerned with human nature and with nature itself that causes us to belief in the necessity of causal connections:

> "Nature breaks the force of all sceptical arguments in time and keeps them from having any considerable influence on the understanding."

Hence it seems that Hume was convinced of the necessity of causal relations, but nevertheless he could not give rational arguments that induction can justify this conviction. Hume's remarks concerning human nature and nature at all do not present a commonly accepted justification of the method of induction and thus of the necessity of causal relations. For this reasons, other philosophers did not consider Hume's arguments as a proof or a justification of necessary causal connections between several events.

In this sense, Kant writes in the introduction of the Prolegomena:[10]

> "Hume ging hauptsächlich von einem einzigen, aber wichtigen Begriffe der Metaphysik, nämlich dem der Verknüpfung von Ursache und Wirkung [....] aus, und forderte die Vernunft, die da vorgibt, ihn in ihrem Schoße erzeugt zu haben, auf, ihm Rede und Antwort zu geben, mit welchem Rechte sie sich denkt: dass etwas so beschaffen sein könnte, dass, wenn es gesetzt ist, dadurch auch etwas anderes notwendig gesetzt werden müsse; denn das sagt der Begriff der Ursache aus."

The problem that was formulated by Hume is quite clear: In the laws of nature there are causal connections which hold necessarily. Since we observe merely finite regularities we justify these causal connections by induction. We are led to this way of justification by human nature, by custom, and by habit. This is, however – and Hume is completely aware of it – no justification in the sense of rationalism. Hence, how causality can strictly be justified, remains an open question.

9.1.7 Kant

The challenge formulated by Hume was taken up by Kant in the *Critique of Pure Reason*. Kant agrees with Hume, that the necessity of causation cannot

[10] I. Kant (1783, PzM), A7.

be justified strictly by induction and he presents a new completely different way of reasoning. The cognition of the real world is a complex process that consists of several clearly distinguished steps. In a first step we have perceptions, impressions, and sensations in space and time – but without objective temporal order and without any law-like connection. In terms of natural sciences one would think of light spots, sounds, colours, and other qualities. In a second step these appearances are ordered and interpreted by conceptual tools, the categories of substance and causality, in order to obtain objective knowledge about the real world. Objective means in this context that the cognition refers to the external, material world and not – or not only to the subjective impressions of the observing subject.[11]

The application of the category of substance to the appearances has to confirm that the changes of appearances in time are merely alterations of predicates which refer to a substance which abides. Hence, according to Kant, we have to find out

> "that all change (succession) of appearances is merely alteration".[12]

In other words, the interpretation of changing appearances as alterations of predicates presupposes a time independent carrier of properties corresponding to these predicates. Hence, the persistence of a substance, a carrier of properties, is not an empirical matter of fact but a necessary precondition of the experience of objects.

> "Permanence is thus a necessary condition under which alone appearances are determinable as things or objects in a possible experience."[13]

On the basis of this clarification we can apply the category of causality. We perceive that predicates or properties change in time, and we try to connect these perceptions at different instances of time such that they refer to an object that persists in time. Here, however, we are confronted with the following difficulty: Time cannot directly be observed such that the perceptions can be ordered according to their temporal sequence.

> "Time cannot be perceived in itself and what precedes and what follows cannot, therefore, by relation to it, be empirically determined in the object."[14]

In order to obtain *objective* knowledge of an object of experience the temporal sequence of various perceptions must be determined. This can be performed by means of the category of causality: *If* two successive appearances may be considered as cause and effect, *then* the cause event is earlier than the effect

[11] For the constitution of objects within the framework of Kant's philosophy see also Sect. 10.1.3.
[12] Kant (1787, KRV), B233.
[13] Kant, ibid. B232.
[14] Kant, ibid. B234.

event. Hence the chronological order of two appearances is not determined by time, which is not observable, but the two events determine their temporal order by causal relations which connect them.[15]

A specific difficulty of understanding Kant's arguments comes from the fact that this way of reasoning can be inverted and is usually applied by Kant in its inverted form – which inversion is also called the transcendental way of reasoning. Whenever we are given objective experience of some object, then this object must have been constituted on the basis of appearances in space and time by means of the categories of substance and causality. Hence we arrive at the result that all alterations which refer to properties of an object are in accordance with the law of causality. In other words, the law of causality holds necessarily, since causality belongs to the preconditions of any objective experience. In this sense, Kant considers the causality law as a law that holds a priori in our experience.

This result establishes the principle of causality which is expressed by Kant as follows:

> "All alterations take place in conformity with the law of the connection of cause and effect."[16]

Even more instructive is presumably the formulation of this principle in the first edition of the *Critique of Pure Reason*:[17]

> "Everything that happens, that is, begins to be, presupposes something upon which it follows according to a rule."[18]

In this formulation we find also an interesting link to the concept of a "law of nature". The "rule" which is not further specified here, may be considered and understood as a law of nature. The principle of causality does not claim that a particular law of nature holds a priori, it merely states that under the assumed premises *there is* a law that connects the two events in question.

9.1.8 von Helmholtz

Herrmann von Helmholtz (1821–1894) was one of the leading scientists of the 19th century. He worked in physiology, physics, and geometry and he was

[15] It should be emphasised that Kant did not assume that any two appearances can be chronologically ordered according to this rule. However, if this were not the case, we would not obtain objective knowledge: *"If each representation were completely foreign to every other, . . . no such thing as knowledge would ever arise."* (Kant, ibid. A 97).

[16] Kant, ibid. B232.

[17] Kant, ibid. A 189.

[18] On account of the importance of this principle for Kant' transcendental way of reasoning we add here the original German version: *"Alles, was geschieht (anhebt zu sein) setzt etwas voraus, worauf es nach einer Regel folgt."*

actively involved in the epistemological discussion of his time. Although he assumed an empiricist position he discussed Kant's transcendental arguments very carefully. His main critique against the Kantian approach was based on his physiological investigations which had led him to the conviction that the form of our spatial intuition is not given a priori – as it was assumed by Kant – but determined by the physiological possibilities of our sensual perceptions.[19] On the basis of this result Helmholtz concluded that Kant's "proof" of the Euclidean character of the geometry is not tenable and must be replaced by a relaxed argument which refers to the material preconditions of our spatial intuition and which allows for the three non-Euclidean geometries.[20] This way of reasoning could be justified mathematically by Helmholtz' two theorems on non-Euclidean and Riemannian geometry.[21] With respect to the causality law Helmholtz' position is somewhat nearer to Kant's transcendental way of reasoning. However, he shares Hume's scepticism against the justification of induction:

> "Jeder Induktionsschluß stützt sich auf das Vertrauen, dass ein bisher beobachtetes gesetzliches Verhalten sich auch in allen noch nicht zur Beobachtung gekommenen Fällen bewähren werde."[22]

Nevertheless, he accepts Kant's argument that our sensual perceptions must be ordered and interpreted by conceptual prescriptions – which Kant called *categories*. However, Helmholtz is aware of the fact that the grasp of material and external reality is not necessarily successful. It is based on the trust in the law-like behaviour of the real world. The assumption that for an observed process a cause can be found, is a *regulative principle* that is called "law of causality". It is obvious that this "causality law" does not hold necessarily.

> Für die Anwendbarkeit des Causalgesetzes haben wir aber keine weitere Bürgschaft als seinen Erfolg. Wir könnten in einer Welt leben, in der jedes Atom von jedem anderen verschieden wäre, und wo es nichts Ruhendes gäbe. Da würde keinerlei Regelmäßigkeit zu finden sein, und unsere Denktätigkeit müsste ruhen.[23]

In this sense – and not in the Kantian one – Helmholtz says

> "Das Kausalgesetz ist wirklich ein a priori gegebenes, ein transcendentales Gesetz."

This interpretation of the law of causality as a methodological principle, which allows for recognising the real world, is more tolerant than Kant's interpretation of causality as a necessary precondition of any possible experience. According to Helmholtz the causality law is not given priori to any experience,

[19] Helmholtz (1879 TdW), pp. 27, 34, and 48.
[20] Helmholtz l.c., pp. 28, 29.
[21] Helmholtz (1884 UBA) and (1868 TdG).
[22] Helmholtz (1979, TdW). p. 46.
[23] Helmholtz l.c. p. 47.

since its applicability depends on its success. Helmholtz' position anticipates in some sense the critique of Kant's rigorous a priorism that was expressed by physicists in the 20th century. In special relativity and in quantum mechanics only a relaxed and more general concept of causality can be applied to the physical phenomena.[24]

9.2 Proposed Answer – General Part: Properties of the Causal Relation

General Remarks

All laws of nature represent some causal relation, provided the two following conditions are satisfied:

(i) "Causal relation" is defined differently for dynamical and for statistical laws on the one hand and also for different areas of physics like classical mechanics, theory of relativity, thermodynamics and quantum mechanics on the other.
(ii) In all cases of application, the causal relation is understood in such a way that it satisfies the *chronology condition* (i.e. there are no closed time-like curves) and the condition of *temporal order* (the cause preceeds the effect).[25]

Concerning (i) it will be shown that the properties of the causal relation differ not only for different types of laws, but also for different applications of one type of law in different areas. This indicates already that we do not propose one single concept of causality with global applicability, although certain features of the causal relation will appear in many versions, even if they are sometimes only partially satisfied.

From (ii) it follows that only those laws of nature which describe a time development of a physical system represent such a causal relation, which will be described in a more detailed and precise way subsequently. This does not mean, however, that laws of nature, where (ii) is not satisfied or not applicable, like the ideal gas law $pV = kT$ (because of its timeless formulation), describe physical systems in which no causal relations would exist, but only that they must be of an entirely different type than the ones which satisfy (ii) and are represented by laws of nature.

From what has been said already it is clear that we do not give a treatment of all types of causal relations. The focus here is only those causal relations that are represented by laws of nature. And, since causal relations described by laws of nature occur either always (dynamical laws) or in most cases (statistical laws), it follows, as a further restriction, that we do not treat causality

[24] cf. Mittelstaedt, P. (1989, PMP), Chaps. I, IV, and V.
[25] For these properties recall 7.2.3.5(2b) above, and for more details see Sect. 9.3 (Proposed Answer, Special Part) below.

as a relation between single events which happen only accidentally. Thus, Kutschera,[26] for example, concentrates on single causes and single effects by emphasising that there need not be any lawlike connection, like in the example where somebody breaks his leg because of a banana skin. Investigations on single, apparently non-law-like causal relations have their own right, and in fact, there are probably laws involved also in such cases. What we want to point out here, however, is to say that they (and other causal relations which are not law-like) are not a topic of this chapter (or book). Another approach, which does not concern us here is a kind of ontological foundation of causal relations by a theory of propensities or capacities underlying all single causes.[27] Such proposals go back ultimately to the theory of potency of Aristotle. Again, such investigations have their own right. But since the topics of this book are *laws of nature*, this is not an adequate place for this kind of ontological investigations. The subsequent parts of 9.2 are divided as follows: First, properties of the causal relation will be discussed in 9.2.1–9.2.2 with respect to the question which of them are present in the causal relation represented by laws of nature. Second, two recent views (regularity and counterfactuality) about causality will be discussed (9.2.3). Third, a proposal will be made for a principle of causality both for dynamical laws and for statistical laws (9.2.4).

We shall divide the properties of the causal relation into three parts: (1) Logical properties: Irreflexivity, asymmetry, transitivity; one-to-one, one-to-many, many-to-one (Sect. 9.2.1). (2) Spatio-temporal properties: Continuity, temporal order, limitation of causal propagation, objectivity of causal ordering. (These will be treated in detail in the chapter on causality in different areas of physics (9.3).) (3) Intrinsic properties: Completeness, robustness, necessity (Sect. 9.2.2).

9.2.1 Logical Properties

(1) Irreflexivity

Are all causal relations which are represented by laws of nature irreflexive? If $C(s_1,\, s_2)$ stands for "state s_1 causes state s_2", then the causal relation C, represented by the law, is irreflexive if and only if $\neg(\exists_x C(x,x))$. Thus, a cause cannot be identical with the effect of which it is the cause, although in a causal chain a state s_2 which is an effect relative to the ancestor state s_1 may be a cause relative to its successor state s_3. One may think that there are exceptions with periodic systems, i.e. systems in which the state of the system recurs after a finite period of time. On a closer look, however, the repeated state is not identical with the earlier; first, since it occurs at a different time, and second and more importantly, since,

[26] cf. Kutschera (1993, Caus), p. 569.

[27] Recent proposals of that sort have been made by Popper (1959, PIP), Suppes (1970, PTC), and Cartwright, (1989, NCM).

as Bohm expresses it, "there is no known case of a causal law that is completely free from dependence on contingencies that are introduced from the outside of the context treated by the law in question."[28] That means to say, that the respective state "repeats" itself is an idealisation abstracting from deviations coming from the contingencies mentioned in the quotation above. It will be clear that we need not to mention statistical laws, where the subsequent microstates of the system always differ from one another.

Historically, Thomas Aquinas required irreflexivity for the causal relation but not asymmetry and not transitivity (recall 9.1.2 above). The concept of "causa sui" (cause of itself) originated much later, probably with Spinoza's Definition 1 of his *Ethics*:

> "I understand that to be a cause of itself the essence of which involves existence"[29]

From the requirement of irreflexivity for causality, assumed by the most important philosophers and theologians in the Middle Ages, it follows that they understood causality as a relation either among events of this universe (or of creation) or between God and his creation, but not with regard to God himself. For our purposes, causality is understood as a relation among states of this universe. But it seems to make no sense at all to say that a state (of this universe) causes itself or is the cause of itself.

From the historical Sect. 9.1.1 it is also clear that Aristotle assumed irreflexivity for the causal relation, since he assumed even the something stronger asymmetry from which irreflexivity follows. Also Leibniz's principle of sufficient reason presupposes that the sufficient reason (cause) for an event is not identical with that event. Moreover, Newton's causes as forces are not identical with their effects (like acceleration). Further, we see that in the interpretation of the dynamical law as a causal law by Lagrange, Laplace, Hamilton, and Maxwell the initial state as the cause is always different from the final state as the effect. Also with respect to statistical laws the states that can be called causes (a microstate at a certain time or a series of microstates in a time interval) and those which can be called effects (a later microstate or a series of microstates at a later interval of time) are always different.

It seems reasonable, therefore, to postulate irreflexivity for all types of causal relations (represented by laws of nature), which will be treated here subsequently. We may sum up the reasons for not accepting causal relations of the form "x causes x" as follows: *First*, because our common

[28] Bohm (1957, CCM), p. 61.

[29] Spinoza (1677, Eth) I, Def. 1. The original text in Latin reads: "Per causam sui intelligo id, cujus essentia involvit existentiam; ...". Whether such a conception is reasonable at all is questionable. But this is not the place to enter such a discussion.

understanding of cause and effect presupposes that they are not identical. *Second*, because the understanding of cause and effect which is in accordance with the philosophical and scientific tradition also tells us (only with very few exceptions, like Spinoza) that the causal relation has been interpreted as irreflexive. *Third*, because there seem to be no physical models with causes and effects as physical states where the causal relation is reflexive.

(2) Asymmetry

According to the terminology introduced at the beginning of Chap. 5, a binary relation (here, the causal relation) is symmetric if and only if

$$\forall_{xy}(C(x,y) \rightarrow C(y,x))$$

and asymmetric if and only if

$$\forall_{xy}(C(x,y) \rightarrow \neg C(y,x)) .$$

We call it non-symmetric if and only if it is neither symmetric nor asymmetric.[30]

A further remark is necessary: The causal relation is understood, here, as a binary relation $C(\mathrm{S}_1, \mathrm{S}_2)$ between two states S_1 and S_2. However, we have to add the specification: ... between two states S_1 and S_2 that are connected by laws of nature. But it can be understood also as a relation between a property A_1 of state S_1 and a property A_2 of state S_2, which are connected by laws of nature.

After this clarification, we come back to the question of whether the causality relation expressed by laws of nature is always asymmetric, or whether there are also cases of symmetry. On a first impression, it seems that the following examples are cases of a symmetric causal relation expressed by laws:

- the causal relation expressed by Newton's law of the mutual gravitational interaction between two bodies;
- the causal relation expressed by Coulomb's law of the mutual electrostatic interaction between two charged bodies;
- the mutual interaction between two neurons firing to each other.

At first glance, the two physical examples mentioned, the gravitational interaction (according to Newton's law) and the electrostatic interaction (according to Coulomb's law), give in fact the wrong impression of a symmetric causality relation that is induced by an instantaneous *action at a distance*. However, on a closer look the two examples are not cases of a symmetric causal relation. They could be interpreted as showing a symmetric causal relation only if we make the additional assumption that there is an instantaneous *action at a distance* that can be interpreted as

[30] It should be mentioned that sometimes a different terminology is taken, but we adopted that which is widely used in logic textbooks.

a causal interaction. In accordance with contemporary physics we can, however, avoid this conclusion, if instead of an instantaneous interaction we make use of the concept of a dynamical *field* that propagates in space-time with a finite velocity, and which was introduced first by Faraday and Maxwell into electrodynamics. In addition, a field of this kind that propagates with a finite velocity $v \leq c$ fulfils also important requirements of special relativity.[31]

That under these assumptions the causal relation in the examples mentioned cannot be symmetric can be further illustrated by the diagram in Fig. 9.1.

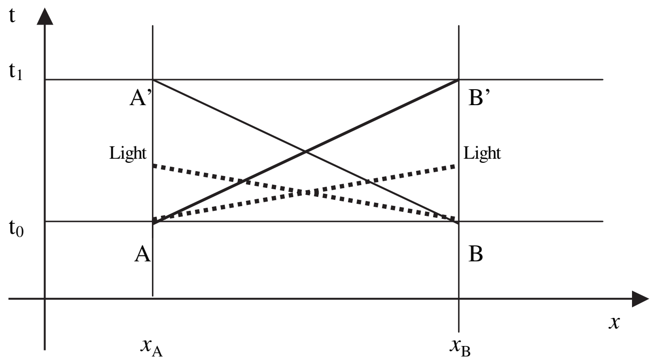


Fig. 9.1. Mutual causal influence between two physical systems

Let A and B be two different physical systems at rest in an inertial system $I(x,t)$. The two systems are at different places x_A and x_B, respectively, but at the same time t_0. Light rays emitted by systems A and B in opposite directions meet half way between the systems at time $t_L = t_0 + (x_B - x_A)/2c$. A causal influence from A to B does not reach B at t_0 but at a later time $t_1 = t_0 + \Delta t$, where Δt is the time needed for the propagation of the causal influence. According to special relativity, we have $\Delta t \geq 2(t_L - t_0)$.

But B at t_0 has meanwhile developed into B$'$ at t_1, and strictly speaking B $\neq$ B$'$. The same holds for the causal influence from B at t_0. It cannot reach A at t_0 but reaches A$'$ at t_1 (into which A was developed during the time interval Δt). And again, strictly speaking, A $\neq$ A$'$. Hence, the

[31] It is, of course, somewhat surprising that by means of the static gravitational potential the two-body problem in celestial mechanics can be calculated with a very high degree of accuracy. If we discuss the analogous problem with two charged bodies and the Coulomb potential, it becomes obvious that the electromagnetic field must be taken into account, since otherwise the radiation emitted from the moving bodies could not be explained. However, the difference between the two cases is only a quantitative one. The ratio of the coupling constants is about 10^{40}, and hence it is very hard to measure gravitational waves.

above examples for causal interactions are not cases of symmetric causal relations expressed by laws, which describe these interactions.[32]

If A causally influences B′ (where the influence is described by a law), then it is not the case that B′ causally influences A. Therefore, we have to speak of an asymmetric causal relation. Observe, however, that the asymmetry is not due to a different causal influence; on the contrary, the kind of mutual causal influence (say electrodynamical interaction, gravitation, neuronal firing) can be the same (and it is the same in the above cases), but, since the causal influence cannot be instantaneous, the causal relation cannot be symmetric. In other words: Since causal relations expressed by laws of nature have to obey certain necessary preconditions, they cannot be symmetric. Here, the important preconditions are: Temporal order (the cause precedes the effect), limitation of causal propagation (the causal propagation has a finite limit), and, implicitly, also the chronology condition (there are no closed timelike curves).[33]

If we drop the condition of temporal order, the causal order would not imply time order (from past to future) even if it could still imply time orientation, i.e. the distinction between past and future. In this sense, all the dynamical laws that are compatible with time reversal (like the laws of classical mechanics) do not presuppose temporal order. Thus, state S_1 at t_1 can be the cause for state S_2 at t_2 just as S_2 at t_2 (the "effect") can be the cause of S_1 at t_1, where the causal connection is described by a dynamical law. In such a situation we could say that the causal relation is symmetric. However, our understanding of the causal relation both in philosophy and also in natural science is such that we presuppose temporal order as a necessary condition for every causal relation. Therefore, in the case of a dynamical law of classical mechanics all the past events can count as causes for the future events, but not vice versa (cf. 9.3.1). Under this assumption, the causal relations expressed by dynamical laws are asymmetric.

On the other hand, there are many examples of causal interactions described by statistical laws where the causal relation is asymmetric: The positive energy of state S_1 causes the positive energy of state S_2, where the respective law describes the propagation of energy from state S_1 to state S_2. But the positive energy of S_2 is not the cause of the positive energy of S_1. The past microstates $MI_i, \ldots, MI_n$ (of a gas, say a litre of air at 273 K) cause the microstate MI_{n+1}, but this microstate does not necessarily cause one of the past microstates, but may cause another possible microstate which did not occur so far.

[32] Another example, which also shows that causal propagation needs time is discussed by Schurz (2001, CAI), p. 60f. cf. further the example discussed by Hausman (1998, CAS), p. 44f.

[33] Concerning the application of these conditions to physical theories see Sect. 9.3 below.

(3) Transitivity
The causal relation between events (or states) x, y, z is transitive if and only if $\forall_{x,y,z}\left[(C(x,y) \wedge C(y,z)) \rightarrow C(x,z)\right]$. We first observe that the causal relation taken in general is not always transitive. For example, it strikes one's eyes that the human genealogical tree is, of course, a causal chain, but not transitive, provided cause is understood as sufficient condition, i.e. as sufficient for producing the effect:[34] The four grandparents are a sufficient cause for the parents and the parents are a sufficient cause for their children, but the grandparents are not a sufficient cause of their grandchildren. In the case of Laplace, a cause (any earlier state) is always both sufficient and necessary. If *cause*, however, is interpreted as necessary condition, then the human genealogical tree is transitive.

The general question whether there is a counterexample against the transitivity of the causal relation can be brought into the following simple form. Is there a case which can be described as follows: State A is capable of changing state B and state B is capable of changing state C; yet state A is incapable of changing state C.[35]

Now, even granted that the causal relation in general is not transitive, the concern here is the causal relation as it is expressed by laws of nature. And, concerning this specification, we can see an important difference with respect to both types of laws, dynamical laws and statistical laws. In the case of dynamical laws of classical mechanics, in the sense of Laplace (cf. 7.2.1.2 and 9.4.3.1) there is transitivity of the causal relations between the states of the system not only in the direction towards the future, but also (because of the time-reversal invariance of the laws) in the direction towards the past. This is so also concerning the causal relation expressed in Maxwell's equation in accordance with special relativity. Moreover, the Minkowski spacetime of special relativity allows defining a causal relation that is irreflexive, asymmetric, and transitive.[36] Thus, in the case of dynamical laws, transitivity of the causal relation expressed by these laws is satisfied.

[34] It is clear that "sufficient condition" has to be taken also under normal environment contingencies. Otherwise a cause could be never sufficient except it is the set of all past events.

[35] This formulation is due to Pearl, who discusses several models which are counterexamples to transitivity and writes: "That causal dependence is not transitive is clear . . . The question naturally arises as to why transitivity is so often conceived of as an inherent property of causal dependence . . . ". Pearl (2000, CMR), p. 237. See also Galles, Pearl (1997, ACR).Transitivity of the causal relation was recently defended also by David Lewis (2000, CIn). However, an attack of one type of counterexample against transitivity (even if it would be successful) (p. 194) cannot prove transitivity of causation as a general property of causation even if it holds in wide areas.

[36] See below 9.3.1(4).

In the case of *statistical laws*, however, there are counterexamples against transitivity or difficulties for its applicability. Consider the following thermodynamic processes in non-equilibrium systems, where causes are understood as sufficient conditions for the respective effects: The earth, embedded in the system sun–earth–cosmic environment, and any living system in its environment are thermodynamically open. It receives high grade energy, and by passing it through it delivers low grade energy to the environment (recall Sect. 7.2.3.4.3(2d)). Thus, the sun, providing the electromagnetic high grade energy (A), causes the state of order and information of the earth (B), and B produces (causes) the low grade energy C. But it does not hold that A causes C. On the contrary, A caused the production of order and information and not its opposite. In a similar way, transitivity is not satisfied in living systems. High grade energy, which they take from nutrition (A), causes order and orthogenesis in the living organism (B), and B produces low grade energy which is delivered to the environment (C). But A is not sufficient to produce C.

Summing up, we may say that counterexamples against transitivity of the causal relation, where cause is understood as sufficient condition, are available in the case of statistical laws applied to thermodynamic systems of non-equilibrium, and therefore are also available in areas which are based on them, like biological systems.

(4) One-to-one, one-to-many, many-to-one

Assume that the development from an initial state S_1 to some successor state S_2 of a physical system is described by some law. Then, the relation between S_1 and S_2 can be one-to-one, one-to-many or many-to-one. If S_1 can be interpreted as a cause and S_2 as its effect, the causal relation described by the law can be also of one of these three types. Now, the nearest case of a one-to-one relation is realised by an isolated mechanical system obeying Newton's laws of motion. It is an idealisation, which is expressed in Laplace's idea.[37] According to Laplace's idea, one arbitrary state allows the calculation (via laws) of any state in the past or in the future. In this case, the causal relation would be symmetric since "time reversal", or better, reversal of motion is possible. Or, to put it into other words: the necessary precondition for causal relations – temporal order – is not satisfied. Since in the case of laws which allow time reversal the causes (initial states) and their effects (final states) can be exchanged.

On the other hand, in all cases where statistical laws are involved, we have the relations one-to-many and many-to-one between the initial states and the (not necessarily immediate) successor states. Examples for relations one-to-many are all cases where the effect (successor state) remains within certain bounds but has a range of indetermination (or error). This is even so with a relatively precise gun, the more with a particle accelerator

[37] cf. 7.2.1.2 and 9.3.1. Concerning the question of the realisation of isolated (idealised) systems see ch. 11.2 on reliability.

or with processes of radiation. Thermodynamic processes have both one-to-many and many-to-one relations among their states: Thus, on the one hand, one microstate develops into many different successor microstates, and on the other hand, many different states lead to an equilibrium. The first shows that, given one microstate as a cause, the effect needs not be unique, i.e. is not necessary. The second shows that for a unique effect the cause needs not be unique, i.e. is not a necessary condition.

The spatio-temporal properties of the causal relation (expressed by laws of nature) will be treated in Sect. 9.3 (Causality in different areas of physics). Since we already need to refer to the spatio-temporal properties in the subsequent sections, we list them here as follows:

(1) Continuity: Is every causal relation (expressed by laws) continuous, i.e. if A is the cause of B, is there always a C between A and B such that A causes C and C causes B? Or: Does every discontinuity imply non-causality?
(2) Chronology condition: There are no closed time-like curves.[38]
(3) Temporal order: Can the effect precede its cause, or is it a general postulate that the cause must be earlier or simultaneous with its effect?
(4) Limitation for causal propagation: Is the velocity of light or another finite velocity the maximal velocity for causal propagation?
(5) Objectivity of causal ordering: Is the causal order with respect to what is cause and what is effect independent of (a) observer and (b) reference system?

9.2.2 Intrinsic Properties

9.2.2.1 Completeness of Causes

The topic Completeness of Laws will be treated in detail in Sect. 11.1 below. Here, the question of completeness will be discussed only with respect to causes expressed by laws. Historically, Leibniz proposed a universal completeness principle with his principle of sufficient reason: Nothing happens without a sufficient reason. Or, as he adds immediately: Every truth has its a priori proof.[39] The first part can also be paraphrased: Nothing happens without a sufficient cause. The idea of completeness behind these principles can, then, be expressed also thus: A set of reasons (causes, states) C is called complete with respect to an area G if every state E in G can be explained sufficiently by some causes c of C (or, if it is determined sufficiently by some c of C). The version with "explained" is an epistemological one, the one with "determined" an ontological one. The above definition leaves open whether there is

[38] See Sect. 7.2.3.5(2b) above.

[39] For a discussion of several versions of this principle see 11.1.3.1 below, and Weingartner (1983, IMS).

one unique cause (in this case, the set C has to be replaced by its unit set) or many causes as elements of C. It also leaves open whether one of the many causes is already sufficient for the effect, or whether they are sufficient only together. Further, it leaves open whether there is only one unique effect (state), or whether there are many. If we want to make a concrete application – say G is the area of classical mechanics – then we see immediately that something essential is missing here. The expressions "can be explained sufficiently" and "is determined sufficiently" are obscure in a double sense: first, they are imprecise, and second, they seem to presuppose a hidden connection between cause and effect. In other words, they do not spell out an important point: the laws that are needed for the connection. Thus, a more precise reading with a concrete application is this:

A set of initial states C (interpreted as a set of causes) at t_1 is complete with respect to the area of Classical Mechanics if every state e of E at $t_2(\geq t_1)$ is uniquely determined by C and the laws of Classical Mechanics (cf. 9.3.1 below). If we replace the reasons, causes or events by those propositions which describe them, then we arrive at the second version of Leibniz's principle above: Every true proposition describing an event (state) in G can be proved a priori from the axioms (and definitions) which describe the reasons or causes. More accurately and applied: A set of propositions C^* describing the initial states (causes) c of C at t_1 is complete with respect to the area of classical mechanics, if and only if every proposition e^* of E^* describing a state at $t_2(\geq t_1)$ follows from C^* plus the laws of classical mechanics. Now, as it is clear from 11.1.3.1 below, Leibniz did not claim that such completeness results are available for every area G. He thought they could be achieved by man if G is the area of logic or of mathematics or of metaphysics (if built up axiomatically).[40] However, he pointed out that such completeness results are not available for man if G is the area of natural science or ethics and jurisprudence, since contingent propositions (viz. contingent events and causes) are involved in the proof process (viz. in the process of events).[41]

These more philosophical ideas became more concrete and applicable to physics by Laplace's idea of the development of the universe, expressed in his essay about probability.[42] In this passage, a very strong completeness thesis with respect to all events (states) of the whole universe is claimed:

T1: One (arbitrarily chosen) state of the (whole) universe (at a certain point of time) plus the laws of nature are sufficiently complete in order to calculate every other state of the universe.

[40] Today we know that he was right with respect to logic if we take first order predicate logic, but not if we take higher order logic, and not with respect to mathematics.

[41] Although ultimately – and for God – there is completeness also in these areas, according to Leibniz.

[42] Recall the quotation in 7.2.1.2, the conditions of dynamical laws, especially D1 in 7.2.3.2 above and 9.3.1 below.

According to thesis T1, the state (say A) which is arbitrarily chosen in order to calculate state B with the help of laws might occur also later than state B. That means that T1 is independent of temporal order, or temporal order is not presupposed in T1. However, the time axis seemed to be understood by Laplace in such a way that the chronology condition is satisfied. A causal interpretation of T1 can be given by a weaker thesis T2 which follows from T1:

T2: One state A of the (whole) universe at time t_1 – where A is interpreted as a cause – together with the laws of nature are sufficiently complete in the defined sense for producing at least one other state B of the universe at the later time t_2 – where B is interpreted as the effect of A.

This weaker thesis T2, in order to have a reasonable causal interpretation, has to satisfy the following three conditions: the chronology condition (cf. 7.2.3.5(2b)), the condition of temporal order, and the condition of limitation of causal propagation. That the chronology condition is important for every causality condition is especially stressed by Hawking and Ellis after they prove a proposition which shows that the chronology condition is a necessary condition for the causal structure:

> "This shows that in physically realistic solutions the causality and chronology conditions are equivalent."[43]

To this view, we want to emphasise five things:

(i) The chronology condition cannot be proved from a law or from generally accepted axioms; it has to be assumed as a precondition or as an axiom.
(ii) One of the main purposes to assume it is the restriction of the number of possible solutions of Einstein's field equations.
(iii) There are many investigations (for example by Kip Thorne and others[44]) of the consequences, like "time travel" of different sorts, if the chronology condition is violated.
(iv) Since the consequences mentioned in (iii) lead to somewhat paradoxical situations, especially with respect to the causal order, many physicists and philosophers do not deny that the chronology condition is a reasonable assumption.
(v) However one should keep in mind that if the minimum length of all closed time-like curves were extremely long (e.g. greater than the age of the universe) the respective cosmology would behave – for all practical purposes – like an ordinary cosmology with non-periodic time (satisfying the chronology condition).[45]

[43] Hawking, Ellis (1973, LSS), p. 192. For situations in QM (tunnelling processes, measurement processes, and EPR correlations) see 9.3.2c below.
[44] cf. for more details Thorne, K. (1994, BHW). cf. note 87 below.
[45] cf. Barrow, Tipler (1986, ACP), p. 449.

Also the following theses T3–T5 in their causal interpretation (i.e. assuming the three conditions above) follow logically from the strong thesis T1:

T3: The first initial state A of the universe together with the laws of nature are sufficiently complete to cause every later state B (at t_i).
T4: The present state A of the universe together with the laws of nature are sufficiently complete to cause every later state B.
T5: All the past states (including the present state) of the universe, together with the laws of nature, are sufficiently complete to cause every later state B.

Commentary on the Completeness Theses

T1, which expresses the idea of Laplace, makes two assumptions which are necessary conditions for the claim: that the laws of mechanics are the laws of the whole universe, and that the initial states which represent the causes can be described with arbitrary precision. Under these assumptions, an initial state (its exact values for position and momentum) of a part or of the whole universe together with the Hamiltonian differential equation (see 9.3.1a, 2 below) is a complete set of causes for the final state (its exact values for position and momentum). However, since these two necessary conditions cannot be realised globally (for the whole universe), the completeness of the causes cannot be realised globally either. This can be seen as follows: First, the laws of the universe are not exclusively laws of mechanics, not even only dynamical laws, as is evident from the fact that the universe is full of processes of radiation, thermodynamics, and expansion which do not obey dynamical laws. And for statistical laws, conditions D1 and D2 (of dynamical laws cf. 7.2.3.2) do not hold. Second, even granted that the laws are laws of mechanics or, more generally, dynamical laws, it is not possible to describe the initial states that represent the causes with arbitrary precision. And, sometimes small differences in the initial conditions will lead to an exponential separation of originally adjacent points in successor states which cannot be calculated exactly any more. From these considerations it follows that the causes for the final or later states are certainly not complete if the system is the whole universe (according to the idea of Laplace).[46] On the other hand, completeness of the causes for the final state (as the effect) can be approximately achieved if the physical system is a sufficiently isolated mechanical system. For example,

[46] cf. Sect. 11.1 in which Laplace completeness is discussed with respect to laws. In order to avoid misunderstandings, we want to stress that when we speak of the laws of mechanics or those of the universe we mean laws in the sense of L3 (see Chap. 1), not the ideal law L4. Otherwise, the ideal true dynamical laws as hidden structures plus the ideal true states could always be claimed to be complete with respect to other or later states. In any case, completeness cannot be claimed any more if ideal true statistical laws with realistic degrees of freedom (branching) are incorporated.

very exact calculations of a future state of the planetary system show that the degree of completeness of the causes can be considerably high. A good example is the calculation of the orbit of mercury, first by Newton's theory and then by Einstein's Theory of General Relativity. -What has been said for T1 holds also for the weaker theorems T2–T5, if only dynamical laws are involved.

Let us now drop the two assumptions, which have been presupposed for the interpretation of Laplace's idea. Then we have to understand laws of nature as including both dynamical and statistical laws. Since neither condition D1 nor condition D2 (of dynamical laws, cf. 7.2.3.2 and 7.2.3.3) hold for statistical laws, there are degrees of freedom for particles and parts of a physical system and possibilities for branching of the trajectories. And, since we do interpret these degrees of freedom and the possibilities for branching realistically, that is, not epistemically as lack of knowledge plus hidden deterministic parameters, the earlier states plus laws do not uniquely determine the final individual or microstate. Concerning statistical laws, however, we have to add a proviso: it is assumed that the initial state and the final state are of the same category: both are microstates. If, on the other hand, we have the situation of a series of microstates leading to one macro-state, this macro-state (for example a state of equilibrium) can be the unique outcome of a series of microstates such that, in this case, the series of microstates is sufficiently complete (even if there are a lot of degrees of freedom) to produce the macro-state.

Therefore, the earlier states (as causes) plus dynamical and statistical laws are not (in general) sufficiently complete for the final state (as the effect). This holds for all the theses T2–T5.

The thesis T3 deserves special attention. It claims that the initial state (considered as the cause) together with the laws is sufficiently complete for every later state (of the universe). To this claim, we may ask two questions.

(i) Is the initial state uniquely defined – given other facts of the universe?
(ii) In what sense could the initial state be complete?

Concerning (i) we may recall that Thomas Aquinas required in his axiom 3 (cf. 9.1.2 above) that, if something causes something else, then there must be a first cause, i.e. the causal chain (S) must terminate after a finite time or after a finite number of causes. Moreover, in this case the "first cause" does not belong to the universe and is assumed to be independent of the universe in the sense that the causal relation is strictly asymmetric. On the other hand, within the framework of general relativity, a "first cause" of the universe that belongs to the universe itself can be shown not to exist (cf. 9.3.1.c2.γ below). This decides the second question: the initial state cannot be complete because it does not exist as a "first cause".

9.2.2.2 Robustness

The considerations in the chapter on transitivity and on completeness (above) have shown that the causal relation expressed by laws is different for dynamical and for statistical laws. In one case it is transitive, in the other it is not, and in the dynamical case the later state is determined by the earlier state which is not so in the case of statistical laws. The question here is whether the causal relation (expressed by laws) is invariant against disturbances of the state which is interpreted as the cause. More accurately there are two main possibilities: (i) small disturbances of the cause may lead to proportionally small changes – described by linearly increasing functions – of the effect; (ii) small disturbances of the cause may lead to disproportionate big changes of the effect, described by exponentially increasing functions. If in case (i) the kind of causal relation does not change, it is called *robust*. Sometimes robustness is defined only for stochastic or probabilistic causality.[47] But the two possibilities (i) and (ii) concern of course dynamical laws as well. This is clear from the stability condition D4 (Sect. 7.2.3.2). We may therefore call a causality relation *robust* if it is invariant under the condition that D4 is satisfied. On the other hand, we shall not expect that the causality relation will not change if D4 is violated, i.e. if (ii) is satisfied. This leads to the question whether in such a case (violation of D4) we can still speak of a causal relation. Healey and Redhead for example require robustness as a necessary condition for causality.[48] We do not find this reasonable because why should a process with weak perturbation be still causal, whereas one with strong perturbation (chaotic motion) is non-causal. Especially since in the case of *dynamical chaos*, although D4 is violated, the underlying laws are dynamical laws. Although it is evident that the causal relations have to be weaker than in the case of satisfying D4, they need not to disappear completely. The view to accept only a strong concept of causal relation and to deny any causal relation, where the strong causal relation is not satisfied, does not seem to us the right strategy. On the contrary, we accept a pluralism of causal relations and different modes of application in classical mechanics and special relativity, in quantum mechanics, in thermodynamics and areas of statistical laws and finally in processes where D4 is violated.[49]

9.2.2.3 Necessity

It is assumed here that the necessity of the causal relation expressed by laws of nature is nothing else than the necessity of these laws. This can be sub-

[47] Thus for example Healey (1992, CRE), p. 283 defines robustness thus: "If R is a stochastic relation between events a and b, then R is robust just in case $P(a/b)$ (the probability of a given b) is invariant under any sufficiently small disturbance of b.

[48] Ibid., p. 287, Redhead (1987, INR), p. 103.

[49] cf. 9.2.4, 9.3.2(b), and 9.4.3.2 below.

stantiated as follows: First, because the causal relations discussed here are those expressed by laws. Other kinds of causal relations like single event causality or single causal processes based on propensities or capacities are not treated here (cf. 9.4). Second, because the necessity according to which a state $S(t_2)$ follows from a state $S(t_1)(t_1 < t_2)$ with the help of a dynamical law is given by the respective dynamical law. Similarly the necessity according to which a microstate $\mathrm{MI}(t_2)$ follows from one of the microstates $\mathrm{MI}_1(t_i)$, $\mathrm{MI}_2(t_i)$, $\mathrm{MI}_3(t_i)(t_i < t_2)$ with the help of a statistical law is given by the respective statistical law. Third, because the necessity of laws of nature can be defined with the help of the properties of these laws (see below). In this sense the necessity of the causal relation expressed by laws of nature can be reduced to the necessity of the laws themselves. The term "necessary" applied to statements including laws has different meanings. We have to mention at least two meanings which are not concerned with laws of nature: the necessity of the laws of logic and that of the laws of mathematics.

(a) Necessity of the laws of logic
That the laws of logic are necessary is usually expressed with the help of an idea of Leibniz saying that the laws of logic hold in all possible worlds. This idea was semantically clarified and restricted (to "accessible possible worlds") by *Possible World Semantics* in the sense that it answered the question what holds logically in all possible worlds. These are the theorems of PL1 (first order predicate logic with identity) plus some modal propositions depending on the underlying modal system. However, theorems of second order logic or of set theory are not included. Thus *Possible World Semantics* gives a semantic interpretation of what it means to say that a theorem of PL1 is valid, or is logically true, or to say that a theorem of PL1 is logically necessary or holds necessarily. That this kind of *logical necessity* is not the necessity of laws of nature can be seen as follows: First, because according to Chap. 3 laws of logic are not laws of nature and therefore the necessity of laws of logic cannot be the same as the necessity of laws of nature. Second, because laws of logic have nothing to do with a time development. If the antecedent of a logically true implication is referred to an earlier state and the consequent to a later state – as used in a prediction – then this relation of succession is an additional interpretation for the respective application. The same logically true implication can equally be used for a retrodiction (from a present or past event to an earlier past event). Third, because the laws of logic are also completely independent of a causal succession.[50] On the other hand this whole chapter attempts to show how and in what sense laws of nature express causal relations.

[50] Recall however the historical part (Sect. 9.1): According to Aristotle a law-like premise answering a "why" question can be called a *causa finalis* (a final cause). But this kind of meaning – also present in the Middle Ages – where, in an analogous way, premises are called causes w.r.t. the conclusion is not any more used today.

(b) Necessity of the laws of mathematics

Possible World Semantics does not answer the question whether a statement of finite number theory, say $2 + 3 = 5$, is necessarily true or holds in all possible worlds. The deeper reason for that is a basic assumption of PL1, which says that there is at least one element (object) in its universe of discourse. This is adequate for logic,[51] but much too modest for mathematics, where one needs infinitely many elements (objects) in the universe of discourse.

To give a further illustration, the statement: "It is possible that there exist two objects with property F" (say two natural numbers), symbolically: $\Diamond\exists x\exists y\,(Fx \wedge Fy \wedge x \neq y)$ is not provable in any of the modal systems which are semantically interpreted by *Possible World Semantics*; but one would expect that this – even without "$\Diamond$" – should be provable and hence – by the rule of necessitation: If $\vdash p$ then $\vdash\Box p$ – necessary. According to Leibniz also the laws or theorems of mathematics are necessarily true in the sense that they hold in all possible worlds. However, here it is not possible to give a simple interpretation with the help of Possible World Semantics. The deeper reasons for that are the incompleteness and two further important differences w.r.t. logic: (i) The universe of discourse contains infinitely many objects which is made explicit by an axiom of infinity.[52] (ii) Whereas the laws of logic can all be understood as universaliseable w.r.t. to predicates (even if predicates are not quantified because of first order), the axioms of set theory, which underlie many laws of mathematics, make existence assumptions of sets or predicates (relations, functions, etc.) which are type-theoretically of higher order.[53]

That this kind of necessity of the laws and theorems of mathematics is not the necessity of laws of nature can be seen as follows: In Chap. 4 we investigated three different fields of mathematics – arithmetic of natural numbers, geometry of the three-dimensional space, elementary probability theory – w.r.t. the question whether the theorems of these fields are laws of nature. In these three cases it could be shown explicitly that the mathematical laws are valid in nature, but they are not genuine laws of nature, since they can be justified by formal means only without any recourse to experience. Hence, the necessity of both kind of laws cannot be the same. Moreover, as in the case of laws of logic also the laws of arithmetic, geometry, and probability have nothing to do with the time development and are also completely independent of a causal succession.

In the three mathematical fields mentioned the abstract formulation in terms of axioms – e.g. the Peano axioms for arithmetic, the Hilbert axioms

[51] cf. the discussion of Russell in his (1919, IMP), p. 203f. There are weaker logics – so-called "free logics" and "empty logics" – which do not make even this assumption. cf. Bencivenga (1986, FLg). The differences between logic and mathemetics are discussed in Quine (1970, PLg), Boolos (1998, LLL), Sect. 3, Weingartner (1976, WTh II) Sects. 2.21–2.28, and Chap. 3 of this book.

[52] Fraenkel, Bar Hillel, Levy (1973, FST), p. 23ff. and 46f.

[53] cf. Weingartner (1975, FAM) and (1976, WTh II), Sect. 2.21ff.

for geometry, and the Kolmogorov axioms for probability[54] – and its justification by the construction of the mathematical entities, must be clearly distinguished from the validity of the respective theorems in the empirical reality. The demarcation between the two aspects seems to be the explicit reference to possible processes that allow for counting, measuring and constituting the real objects in question. The necessary validity of the mathematical theorems in nature is then a consequence of the necessity of these laws in mathematics. Obviously, this is not the necessity of genuine laws of nature.

(c) Necessity of the laws of nature

The necessity of the causal relation expressed by laws of nature is nothing else than the necessity of these laws. Several reasons for that have been given already above with respect to both dynamical and statistical laws. What concerns the causal relations expressed by laws, we have already described them with the help of logical and intrinsic properties, whereas its spacio temporal properties will be dealt with below. Here the task is to tell what the necessity of a law is. *One* general answer is that the necessity of laws of nature is their invariance.[55] This can be defended as follows:

(i) From a comparison with its opposition: Statements which are not necessary, i.e. statements which describe contingent facts, like initial conditions, do not express an invariance or symmetry. On the other hand the invariance structure "in the laws" implies that they do not hold contingently, but necessarily: "Nevertheless, there is a structure in the laws of nature which we call the laws of invariance."[56]
(ii) From the universal status of the core of a theory which consists of laws: not the contingent facts are contained in the core of a theory, but necessary correlations: "The irrelevant initial conditions must not enter in a relevant fashion into the results of the theory."[57]
(iii) From a definition of natural necessity. Such a definition was proposed by Popper: "A statement may be said to be naturally or physically necessary iff it is deducible from a statement function which is satisfied in all worlds that differ from our world (if at all) only with respect to initial conditions."[58] Popper's definition can be expressed also with the concept

[54] Peano (1889, PAP), Hilbert (1962, GlG), Tarski's axiomatisation and decision method of elementary geometry (1951, DME), Kolmogorov (1956, FTP).
[55] That every law of nature expresses an invariance or symmetry was substantiated in detail in Chap. 5.
[56] Wigner (1967, SRf), p. 29. The special view of Wigner of "laws of invariance" as laws about (or in) laws was discussed in detail in Sect. 5.3.1(3).
[57] Wigner, ibid. p. 8.
[58] Popper (1959, LSD), p. 433. Here "statement" is to be understood as a law statement which is neither a law of logic nor a law of mathematics.

of invariance:[59] A law statement may be said to be naturally or physically necessary if it is invariant under a change of initial conditions.

From Sect. 5.3.2 it is clear that there are further properties of invariance of laws of nature than those w.r.t. initial conditions. Thus we might distinguish more general laws of nature from less general ones by the degree in which they possess invariance properties. And consequently we may say that a law of nature is naturally or physically necessary to a high degree if it possesses invariance properties to a high degree. And since we have said that the necessity of the causal relations expressed by laws of nature is just the necessity of the respective laws, it follows that the causal relation expressed by a law of nature is necessary to that degree to which this law of nature possesses invariance properties.

(iv) From a consideration of the "symmetry group of nature". We understand by "symmetry group of nature" the set of all changes which do not change laws of nature.[60] If we think in terms of models or possible worlds in which the laws are satisfied, then the symmetry group of nature is the set of all models or possible worlds in which the laws of nature are satisfied. In Sect. 8.1.6 it was defended that this set contains more than just one member (i.e. just the actual world). According to these considerations, we can say that the necessity of the causal relations expressed by laws of nature is determined by the set of all models or possible worlds in which these laws are satisfied. Or in other words: the necessity of the causal relations expressed by laws of nature consists in its (the causal relations') invariance w.r.t. a change from one possible world (model) to another.

9.2.3 Two Recent Views: Regularity and Counterfactuality

9.2.3.1 Regularity

The theory of regularity is usually attributed to Hume. Hume assumes that in the complex idea of causation the following elements are present: Temporal order, spatial and temporal contiguity, regularity and necessary connection.[61] Spatio-temporal contiguity is later (in the Enquiry) given up since Hume accepts there causal relations also between mental actions. Necessity is also given up, at least epistemically in the sense that we can only observe finite regularities, the expectations of which are based on our habits and customs. Thus what remains are temporal order and finite regularities. The main reason for

[59] A completely different way of interpreting the necessity of law of nature by means of categories, logical structures and the "law without law" argument is elaborated in detail in Chaps. 10 (categories), 12 (law-without-law) and 13 (logical structures).

[60] See Sect. 5.3.1, note 13.

[61] Hume (1739, THN), Book I, III, Sect. 2 and Part IV, Sect. 5.

Hume that we cannot extend the finite regularities to something more law-like and necessary is the problem of induction: From some observations about finite regularities we cannot derive a necessary causal or law-like relation.[62] In recent literature, theories have been called "regularity theories of causation" which go far beyond Hume's restriction. They propose definitions of *cause* and *principles of causality* which describe the causal relation which is expressed by dynamical laws: For example by saying (roughly) that a cause is an event, which together with laws (dynamical laws are meant) are sufficient for the (event) effect. Or by presupposing the principle of causality: the same cause is always accompanied by the same effect. This principle holds also only for dynamical laws.[63] That such and similar theories cannot count as regularity theories in Hume's sense is seen easily by the fact that Hume's theory lacks a theory of genuine laws (in this case: dynamical laws) expressing a necessary causal relation. Similarly, Mackie's proposal, which is to a great extend John Stuart Mill's, is stronger than Hume's, although it may seem to be a regularity theory in Hume's sense on a first look. But on a closer analysis the claim that a cause is an INUS condition (i.e. an insufficient but non-redundant part of an unnecessary, but sufficient condition) is only applicable under the presupposition that the INUS condition is deterministically law-like. Otherwise it would give us accidental and irrelevant "causes".[64]

Coming back to a regularity theory of causal connection in the more modest sense of Hume we may say the following: The regularity theory of causation is insufficient in the sense that it lacks laws which express causal relations and it is not able to distinguish genuine causes from spurious ones. Therefore an event or state C, which regularly precedes another, E – say a bell of a clock w.r.t. to a bell of another clock ringing later – would be the cause (C) of the event E although in fact it is not. On the other hand it might be suitable for a phenomenological (surface) description for both cases of single event causality and cases for types of causal relations. In this sense a regularity view is always present on a first level of investigation of assumed causal connections which are expressed by new hypotheses which penetrate into an inexperienced area. But if the hypothesis is better and better confirmed with the tendency to become a well-established law the regularity view becomes more and more insufficient. In other words the regularity view is not suitable to interpret the causal relation as it is expressed in laws of nature.

---

[62] For Hume's view on causation see Sect. 9.1.6 and Hausman (1998, CAS), Chap. 3.

[63] See for example Breuer (2001, UCP), p. 163, and Grasshoff, May (2001, CRg), p. 88. We need not to go into further detail here because in this chapter causal relations are treated anyway as they are represented in either dynamical or statistical laws.

[64] cf. the analysis in Hausman (1998, CAS), p. 40ff.

9.2.3.2 Counterfactuality

The counterfactual understanding of the causal relation goes also back at least to Hume. In his Inquiry he says: "We may define a cause to be an object, followed by another... where if the first object had not been, the second had never existed."[65] Recently David Lewis has constructed a theory out of it[66] which enjoys widespread attention. Here we have to consider the question whether the counterfactual theory offers a correct interpretation of one of those types of causal relations which are expressed by laws of nature We shall therefore first give a formulation of the gist of the counterfactual relation as an interpretation of the causal relation and then check whether it suits the causal relations expressed by different types of laws in different areas. The gist of the counterfactual interpretation of the causal relation can be expressed thus:

CF A causes B iff A and B are distinct events and if A were not to occur, then B would not occur either.[67]

From this it follows that *cause* is interpreted as a necessary condition for the effect. Lewis and other authors, in order to evaluate the right part of CF consider possible worlds in which A does not occur, and which of those possible worlds are more similar to the actual world than others. For our consideration it is not necessary to enter such complications. CF and its interpretation of the cause as a necessary condition for the effect is sufficient to answer the question of the applicability to the causal relation expressed by the laws of physics.

(1) If we look first at the laws of classical mechanics, then we come to the following conclusion: Under the assumption that we presuppose temporal order of the causal relation and interpret state A at t_1 as the cause and state B at t_2 ($t_1 \leq t_2$) as the effect, then state A is certainly a necessary condition (which is together with the dynamical deterministic law sufficient) for sate B.[68] However, on a closer look we see that this is only true under the additional assumption that the time development of the system (described by a special solution of the differential equation) begins not later than A. Otherwise any other state which lies between A and B on the time coordinate is equally sufficient together with the law to bring about B. And then A is not necessary for B (or if A were not, B could nevertheless be).[69]

[65] Hume (1748, EHU), Chap. VII, part II.

[66] Lewis (1973, Ctf) and (1973, Caus).

[67] cf. Lewis (1973, Caus), p. 557, and Hausman (1998, CAS), p. 112.

[68] Lewis is aware of the restriction of his analysis to dynamical-deterministic laws (cf. his (1973, Caus), p. 559), although not even this is correct since the counterfactual theory does not apply to dynamical chaos, which is also guided by dynamical-deterministic laws.

[69] In this connection March (1957, NDM), p. 106 f. points to a blind spot, widespread under philosophers which is to single out only one particular state (of a physical

(2) If we consider special relativity then we have to add to the presuppositions temporal order and chronology condition the limitation of causal propagation and the fact that causal propagation needs time. Under these additional assumptions the same can be said about the application of CF as in the case of classical mechanics.
(3) This is however not the same in all cases where dynamical laws are applied. The first deviation is quantum mechanics: In the case of the application of the Schrödinger dynamics the causality relation is not applicable in the sense that a cause or an effect is a completely determined object or state because – in contradistinction to classical mechanics – it possesses at a certain time t only a limited class of properties, which have to be commensurable.[70] The second deviation is the case of dynamical chaos where we have underlying dynamical laws. Here many different possible states $A_1, \ldots, A_n$ and different conditions may lead to the same type of chaotic behaviour. In this case none of the different states or conditions are necessary. This will become clear in a more general way w.r.t. all processes which are describable only by statistical laws.
(4) In the case of statistical laws we have usually a mixture of the relations one-to-many and many-to-one (cf. 9.2.1(4)).

The main point is that many different states may lead to the same equilibrium, or many different microstates (possible causes) may develop into one particular microstate (effect). Thus none of the causes (of the different microstates) is a necessary condition. This shows that the concept of counterfactuality is not suitable in all the cases where statistical laws are essential in physics (i.e. in processes of thermodynamics, radiation, friction, electric transport, measurement processes in QM, etc.). A special simple counterexample against counterfactuality from the area of chemistry is given by Hausman.[71] Recently Lewis defended also transitivity of the causal relation.[72] However, as it was shown in Sect. 9.2.1(3) above the processes in living systems do not obey such causal relations. Or in other words: The statistical laws which can describe such processes do not express a type of causality relation which would satisfy transitivity.

Summing up we may say that Lewis' construction of counterfactuality as an interpretation of the causal relation is applicable only (with the respective provisos) to the causality relation expressed by the dynamical laws of classical

---

system) as the cause of another (later) state as the effect; whereas for a physicist this is completely unjustified since every state in the past light cone can count as a cause for a state (as the effect) in the future light cone. The reason for that narrow-mindedness may be also a confusion with *single event causality* which is not the topic here (cf. 9.2) but which is certainly very important in other domains like that of criminal law.

[70] cf. Sect. 9.3.2 below.

[71] Hausman (1998, CAS), p. 120ff.

[72] Lewis (2000, CIn).

mechanics (with the exception of dynamical chaos) and special relativity. But in all other areas there are either serious restrictions (as concerning dynamical laws in QM) or wrong application or no application at all (as in all the domains of statistical laws).

9.2.4 Principles of Causality

9.2.4.1 Dynamical Laws

In his *Matter and Motion* Maxwell discussed two principles of causality. First he asks whether the principle "the same causes will always produce the same effects" generally holds.[73] Maxwell continues:

> "To make this maxim intelligible we must define what we mean by the same causes and the same effects, since . . . causes and effects cannot be the same in all respects. What is really meant is that if the causes differ only as regards the absolute time or the absolute place[74] at which the event occurs, so likewise will the effects."

A more simple way is to use "states" (states of a physical system at a certain time t) when formulating such principles. Then Maxwell's principle can be formulated in Arthur March's words:

CP1 The same initial state leads – under the same conditions – to the same series of successor states.[75]

CP1 is a principle of causality which is applicable to dynamical laws. This can be seen from the properties of physical systems D1–D4, described by dynamical laws (Sect. 7.2.3.2).[76] Recall the illustration with the film as an interpretation of Laplace's idea: A catalogue card, representing a state of the universe at t_1 is – with the help of the laws of nature – a cause for any catalogue card representing a later state of the universe at t_2. And CP1 says then that two equal catalogue cards (two equal states) will have – according to the laws – the same series of successor catalogue cards (the same successor states).

We may say that CP1 represents a principle of deterministic causality. The main reason for that is that in this case we have a unique state as a cause –

[73] Maxwell (1991, MaM), Sect. 9, p. 13. For Maxwell's second principle of causality see 9.2.4.2.

[74] It should be mentioned however that Maxwell refuses Newton's absolute space and time in Sects. 17 and 18 of the same work: "All our knowledge, both of time and place, is essentially relative." Ibid. p. 12. Recall Sect. 6.2.1(6). There is a footnote on p. 12 by Larmor who interpreted Maxwell (wrongly) along Newton.

[75] March (1957, NDM), p. 14; (1960, PEG), p. 33.

[76] It should be noted that D3 (the condition that the physical system is periodic) is not necessary, although it was satisfied in most cases when dynamical laws have been confirmed by observational results.

corresponding to a unique solution of the differential equation – and a unique state as the effect – again corresponding to a unique solution of the differential equation. In this case conditions D1, D2 and D4 are satisfied.

A principle of causality which received controversial comments was proposed by Pierre Curie.[77] It says roughly that the symmetries of the cause are preserved as the symmetries of its effects and that the asymmetries of the effects can be found in the asymmetries of the cause:

> "Lorsque certaines causes produisent certains effets, les éléments de symétrie des causes doivent se retrouver dans les effets produits.
>
> Losque certains effets révèlent une certaine dissymétrie, cette dissymétrie doit se retrouver dans les causes qui lui ont donné naissance."[78]

This principle of Curie can be interpreted in two ways: (1) In a very general way, (2) in a specific way of application.

(1) In a very general way we may interpret Curie's principle along the considerations made in Sect. 5.3.3(b) in the sense that relative to the symmetric dynamical law the particular state which is predicted as the effect is some symmetry breaking and so is also the initial state as the cause; i.e. from an asymmetric effect and symmetric laws we may conclude asymmetric initial conditions. This fits also very well to what is said by Curie in the summary (p. 414) and what he calls the negative but certain conclusion; while the positive one which claims the transmission of the symmetry of the cause to the symmetry of the effect does not hold with certainty according to him.

(2) In a specific way Pierre Curie investigated the principle in the domain of crystals w.r.t. the preservation of symmetries of crystal structures. This however is a domain much too small to count as evidence for a general principle of causality.[79]

As a concluding remark we may therefore say that Curie's principle can be understood as a causality principle only in the general sense of interpretation (1). And in this sense it is not a new principle and was discussed already in Sect. 5.3.3(b).

---

[77] Curie (1894, SPP).

[78] Ibid. p. 401.

[79] Whyte (1970, PCP) tried to apply Curie's principle to entropy relaxation and frictional/damping processes. We do not want to enter a discussion on the controversy concerning Curie's principle in a more specific interpretation than the first (1) above. Chalmers, (1970, CPr)and Ismael (1997, CPr)defend it, whereas van Fraassen (1989, LaS), p. 240 thinks it is untenable.

9.2.4.2 Dynamical Laws Underlying Chaotic Systems

The second principle of causality which Maxwell discusses in his *Matter and Motion* is the principle: like causes produce like effects. And he says of it:

> "This is only true when small variations in the initial circumstances produce only small variations in the final state of the system. In a great many physical phenomena this condition is satisfied; but there are other cases in which a small initial variation may produce a very great change in the final state of the system"[80]

The passage shows clearly that Maxwell understood already very well the most important necessary condition for dynamical chaos: the sensitive dependence on initial conditions. That means that very small changes in the initial conditions lead to exponentially increasing bifurcations. Not just simple bifurcations which have been known so far as unstable behaviour or as small perturbations, i.e. it is not a case of "perturbative stability".This property of being sensitively dependent on initial conditions is measured by the (positive) Ljapunov exponent.[81] Exponentially increasing factors were well known for a long time as the example for increasing error of Aristotle shows.[82] As Maxwell says, his second principle is only satisfied if small variations in the initial states lead to proportional small variations in the final states, i.e. if condition D4 (Sect. 7.2.3.2) is satisfied. In this case an approximation of principle CP1 can be assumed as a principle of causality. But if D4 is not satisfied, i.e. for cases of chaotic motion, we may formulate the following principle of causality:

CP2 Two similar initial states (i.e. states with a small distance between pairs of two related adjacent points) lead – under a positive Ljapunov exponent[83] – to two separated systems where the Ljapunov exponent (Kolmogorov entropy) measures the average factor by which the distance between the related adjacent points becomes stretched.

Since the Ljapunov exponent (Kolmogorov entropy) measures also at the same time the loss of information about the position of a point (in an interval) and the increasing disorder, we may formulate a second causality principle for dynamical chaos thus:

CP3 Two similar states (i.e. states with a small distance between pairs of two related adjacent points) lead – under a positive Ljapunov exponent

[80] Maxwell (1991, MaM), p. 13.

[81] For details see Schuster (1989, DCh), p. 24ff. and Weingartner (1996, UWT), p. 52ff. See also below Sects. 9.4 and 7.2.3.2 commentary to D3 and D4.

[82] cf. Sect. 7.2.3.2, note 35, and Sect. 11.1.3.5(3).

[83] More generally: The Kolmogorov–Sinai entropy (or for short: Kolmogorov entropy) which is equivalent to the sum of positive Ljapunov exponents for more dimensional maps according to a proof by Pesin. Observe that dynamical chaos is based on trajectories described by a Hamiltonian. Quantum chaos and chaos in large Poincaré systems (cosmology) have different properties. See below, Sect. 9.4.

(Kolmogorov entropy) – to a certain degree of disorder and loss of information which is measured by the Kolmogorov entropy.

So far we have dealt with causality principles for dynamical laws. And this for obvious reasons: For a long time the (dynamical) laws of classical mechanics have been understood as *the* causal laws. From this point of view it was also not realised that D4 is a very important presupposition which has been a hidden assumption for centuries. Not in the sense that one considered the systems free of perturbation; but in the sense that one assumed arbitrary degree of perturbation (i) could be exactly calculated by special mathematical methods and (ii) would lead to the relaxation of the system after some finite time (even if energy supply is not stopped). Both assumptions were wrong. This was realised theoretically already by Hadamard and Poincaré and experimentally only in the second half of the 20th century. As a result we may say that the dynamical laws which describe physical systems satisfying condition D4, express one type of causal relation which is described by the principle CP1 whereas the dynamical laws which underlie those physical systems not satisfying condition D4 express another (weaker) type of causal relation characterised by CP2 and CP3.

9.2.4.3 Statistical Laws

It took a long time until the statistical laws, mainly discovered in the 19th century, were accepted as genuine laws. Many physicists had a hope like the one formulated by Planck: "I believe and hope that a strict mechanical significance can be found for the second law along this path, but the problem is obviously extremely difficult and requires time."[84]

It was understandable therefore that one spoke of causality, if at all, only in connection with dynamical laws. In addition to that the properties of statistical laws made it difficult to find a causal relation which was expressed by such laws. This is also clear from the four properties of systems obeying statistical laws discussed in 7.2.3.3 and their comparison with those holding for dynamical laws (7.2.3.4). From such a comparison one can easily see that a principle of causality in the sense of CP1 is not applicable here: the same initial states lead usually to different successor states; but the different successor states may represent the same statistics. In thermodynamic terms this means that if we start with two equal micro- or macro-states, the two series of successor microstates will be entirely different, but their average values say the velocity distribution or the mean distance between two molecules, etc. will be the same, such that the macro-states will be the same. Such a development is of course only realised, if both systems are interpreted as isolated systems, i.e. if the two series of successor microstates are not influenced by different environments. If this is granted, we can accept the following principle of March as a principle of causality for statistical laws:

[84] Planck in a letter to his friend Leo Graetz. Cited in Kuhn (1978, BBT), p. 27.

CP4 The same initial state may lead to different series of successor states. But those successor states which belong to the same initial state obey the same statistics.[85]

The principle CP4 is independent of whether the statistical process is one-to-many (branching; example: radiation) or many-to-one (running together; example: many states leading to an equilibrium).

Summing up we may say that there is not only one principle of causality which is applicable in a suitable way to both of the different types of laws, dynamical and statistical laws, and to dynamical ones under more liberal conditions, like in the case of violating D4. On the contrary, we have to distinguish different principles of causality and so to accept a kind of pluralism of concepts of causality in physics. This is also supported by the discussion on transitivity and counterfactuality: in some areas (classical mechanics) the causal relation is transitive, and under some provisos, counterfactual in others; where statistical laws have to be applied it is not transitive and not counterfactual.

9.2.5 Answer to the Objections

9.2.5.1 (to 9.1(a)) As has been demonstrated in the Proposed Answer (general part), a pluralism of concepts of causal relation is necessary when they are applied to physical laws and to the physical systems described by them. And therefore it is correct, as pointed out in the objection, that the causal relation represented by dynamical laws cannot be applied to statistical laws. From this however it does not follow that statistical laws do not represent any causal relation. This causal relation has to be weaker than the one expressed by dynamical laws, which is shown by the difference of the two principles of causality CP1 and CP4 and moreover by several other differences concerning the properties of the causal relation (like transitivity, one-to-many and many-to-one). Moreover, remember that also in the case of dynamical chaos, though guided by dynamical laws, the principles of causality have to be weaker (CP2 and CP3).

9.2.5.2 (to 9.1.(b)) The answer to this objection is clear from the section on Asymmetry (9.2.1(2)). Only on a first impression gravitational interaction seems to express a symmetrical causal relation. As soon as we incorporate the fact that any causal propagation needs some finite time, the asymmetry is understandable.

9.3 Proposed Answer – Special Part: Causality in Different Areas of Physics

In contemporary textbooks of physics we can find the concept of causality only very rarely. Causality is not one of the laws, principles, or axioms of physics.

[85] cf. March (1957, NDM), p. 14 and (1960, PEG), p. 37.

Instead, causality appears in physics merely as a general aspect, sometimes as a regulative principle and as a requirement that is based on philosophical arguments. The reason is, that we are confronted in physics with a situation that is completely different from all philosophical positions mentioned in 9.1.(1–8). Primarily, we are concerned with laws of nature and we will not discuss here why these laws are valid. Instead, we have to check whether these physical laws *do* fulfil the requirement of causality, *can* fulfil it, or *must* fulfil it.[86] Hence, we will investigate here several classes of physical laws, in particular

- laws of classical physics (Newtonian mechanics, special relativity, general relativity),
- laws of quantum physics (Schrödinger dynamics, uncertainty, non-objectivity).

We will then check whether the various properties of a causality relation mentioned in 9.2(1–4) actually pertain to the physical laws in question.

9.3.1 Causality in Classical Physics

(a) Classical mechanics

(1) Newtonian formulation In classical mechanics the spacetime behaviour of a point-like body, a particle, with the inertial mass m is – in an inertial system $I(x_k, t)$– governed by the equation of motion

$$m\frac{d^2x_i}{dt^2} = K_i \qquad (i = 1, 2, 3)$$

where x_i are the space coordinates of the mass point in I and t the absolute and universal time coordinate. K_i are the components of an external force. This ordinary second order differential equation determines the trajectory $x_i(t)$ of the body in spacetime, if for a special value $t = t_0$ initial conditions $x_i(t_0) = x_i^0$ and $\dot{x}_i(t_0) = \dot{x}_i^0$ for the function $x_i(t)$ and for the first derivative are given. In other words, if for $t = t_0$ the position $x_i(t_0)$ of the mass point and its velocity $\dot{x}_i(t_0)$ are given, then the entire trajectory is determined by the dynamical law.

In this theory, causality is realised in the following sense: Every effect $\{x_k(t_1), \dot{x}_k(t_1)\}$ at time t_1 has exactly one cause $\{x_k(t_0), \dot{x}_k(t_0)\}$ at time t_0. However, the concepts of cause and effect are somewhat artificial in classical mechanics. The trajectory of a body is completely determined by the equation of motion together with the initial conditions and it is merely a matter of interpretation if we call a point $A = \{x_k(t_0), \dot{x}_k(t_0)\}$ in the configuration space the "cause" of another point $B = \{x_k(t_1), \dot{x}_k(t_1)\}$ at

[86] This point of view is already expressed by Kant's definition of causality: "*Everything that happens, that is, begins to be, presupposes something upon which it follows according to a rule*". – In our case, the "rules" are the laws of physics.

a later time, which we call “effect”. This difficulty can, however, in some sense be eliminated in the Hamiltonian formulation of classical mechanics.

(2) Hamiltonian formulation

In the Hamiltonian formulation of classical mechanics a causality relation can be found which is perhaps somewhat nearer to the intuitive idea of causation meaning that what happens at $(x_k^A,\, t^A)$ causally effects what happens at $(x_k^B,\, t^B)$. Indeed, what happens at $(x_k^B,\, t^B)$ is not “caused” by the spacetime point $(x_k^A,\, t^A)$ alone but also by the forces which act on the particle in question. This additional aspect can be expressed in the following way.

First,we replace the Newtonian equation of motion by the Lagrange equation

$$\frac{\partial L}{\partial x_i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}_i} = 0$$

with initial conditions $x_i(t_0) = x_i^0, \dot{x}_i(t_0) = \dot{x}_i^0$ where $L(x_i, \dot{x}_i, t)$ is a conveniant Lagrange function which contains all information about forces that act on the moving particle. Second, we define new “canonical coordinates” by

$$q_i(t) := x_i(t);\; p_i(t) := \frac{\partial L}{\partial \dot{q}_i}$$

that form the “phase space” $\Gamma(q_k, p_k)$ and define the Hamiltonian $H(q_k, p_k, t)$by

$$H(q_k, p_k; t) = \sum p_k \dot{q}_k - L$$

which contains all dynamical information. The Hamiltonian equations of motion

$$\dot{q}_k = \frac{\partial H}{\partial p_k}, \quad \dot{p}_k = -\frac{\partial H}{\partial q_k}, \quad -\frac{\partial L}{\partial t} = \frac{\partial H}{\partial t}$$

together with the new initial conditions

$$q_k(t_0) = q_k^0\,, \quad p_k(t_0) = p_k^0$$

determine the trajectory $\{q_k(t), p_k(t)\}$ of the particle. Making use of the Poisson operator

$$X(H) = \sum \frac{\partial H}{\partial p_k}\frac{\partial}{\partial q_k} - \frac{\partial H}{\partial q_k}\frac{\partial}{\partial p_k}$$

the Hamiltonian equations of motion read

$$\dot{q}_k = X(H)q_k, \quad \dot{p}_k = X(H)p_k$$

and can be integrated in closed form. Taking together the canonical coordinates q_k and p_k of the phase space we get

$$\{q_k(t); p_k(t)\} = \exp[X(H)(t - t_0)]\,\{q_k(t_0); p_k(t_0)\}\ .$$

This is the most adequate formulation of the determinism of classical mechanics. The phase space operator $\exp[X(H)(t-t_0)]$ causes the phase space point $\{q_k(t_0); p_k(t_0\}$ to move to the phase space point $\{q_k(t); p_k(t)\}$. In this way the particles trajectory is continuously created by the "causality operator" $C^{\mathrm{OP}}(\mathrm{H}) = \exp[X(H)(t-t_0)]$ of classical mechanics.

Wemention briefly some properties of this causality relation contained in classical mechanics that follow immediately from the equations of motion.

(1) Obviously, this causality relation is continuous in the sense mentioned above. If, what happens at $A \equiv \{q_k(t^A),\, p_k(t^A)\}$ can causally effect what happens at $B \equiv \{q_k(t^B),\, p_k(t^B)\}$ with $t^A \le t^B$, then also every $A' \equiv \{q_k(t^{A'}),\, p_k(t^{A'})\}$ with $t^A \le t^{A'} \le t^B$ is a cause of B. The reason is the continuity of the "causality operator" C^{OP}.

(2) The time values t^A and t^B of cause $A(t^A)$ an effect $B(t^B)$ fulfil the relation $t^A \le t^B$, i.e. the cause is earlier than the effect, in accordance with the chronology condition.However, in the present case this order can also be inverted since the event $B(t^B)$ determines also $A(t^A)$ if $t^A \le t^B$. Newtons equation of motion as well as the causality operator C^{OP} are symmetric with respect to time inversion and do not define a certain direction of time.

(3) Since in classical mechanics the existence of an absolute and universal time is presupposed, no problems arise with the objectivity of the causal relation. If two events $A(t^A)$ and $B(t^B)$ with $t^A \le t^B$ are causally connected in one system of reference, then they are also causally connected in any other frame of reference.

(b) Special relativity

(1) The light-cone structure of spacetime

Compared with Newtonian mechanics in special relativity we dispense with the metaphysical concept of absolute time. Spacetime of special relativity is best described by a Minkowski space M, i.e. a four-dimensional pseudo-Euclidean spacetime with signature 2. In the Minkowski spacetime we must presuppose that there are no gravitational fields. The relations between physical laws and causality are more complicated here than in case of Newtonian mechanics but also more interesting. We will briefly describe the new situation.

The four dimensional Minkowski spacetime $(x_k,\, t)$ is characterised by its light cone structure. In any inertial system there is a maximal velocity for the propagation of waves, signals, and in particular for causal chains, which is given by the velocity of light in vacuum.

The boundary of realisable velocities is then given by the light cone $x_k = ct$. Two systems of inertia I and I' are connected by a Lorentz transformation which leaves the light cone invariant. The light cone structure allows for a invariant decomposition of spacetime in three completely distinct regions. In a given inertial system $I(x_k,\, t)$ no effect whatever can

propagate faster than light in vacuum. Hence if the "cause" is at the point A $\{x_k^A, t^A\}$ then a possible "effect" at the point B $\{x_k^B, t^B\}$ can only be lying in the forward light cone of the point A given by

$$L^+(A) = \{\left|x_k - x_k^A\right| \leq c(t - t^A), t \geq t^A\} \ .$$

Inversely, an "effect" at the event point B (x_k^B, t^B) can be effected only by cause events A' lying in the backward light cone of the event B given by (Fig. 9.2)

$$L^-(B) = \{\left|x_k - x_k^B\right| \leq -c(t - t^B), t \leq t^B\} \ .$$

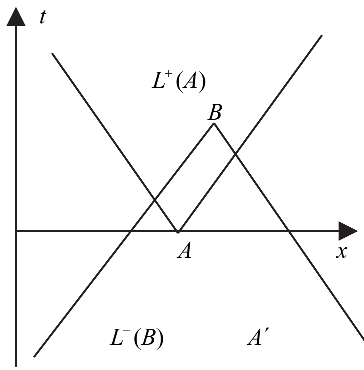


Fig. 9.2. Cause A and forward light cone; effect B and backward light cone

The light-cone structure leads to the following terminological convention: Given an event A, the totality of events B which can be influenced by A, i.e. the events in or on the surface of the forward light cone $L^+(A)$ is called the future of A. The totality of events in or on the surface of the backward light cone $L^-(A)$ is called the past of A. It is obvious that there is a large region in spacetime which belongs neither to the future $L^+(A)$ nor to the past $L^-(A)$ of A. It is called the presence of A. (Fig. 9.3)

Theimportance of this separation of spacetime into future, past, and present becomes obvious if we consider different inertial systems I, I' and I''. An inertial system is a reference system with the special property that all point-like particles which are not influenced by any force, move along straight lines in the sense of Euclidean geometry. Usually, we assume that an inertial system is equipped with clocks and rods and an observer who registers the measured results. Generally, two inertial systems I and I' are in relative motion, where the relative velocity $v_{I,I'}$ is constant in time and always smaller than the velocity of light in vacuum. Two inertial systems I and I' are connected by a Lorentz transformation.

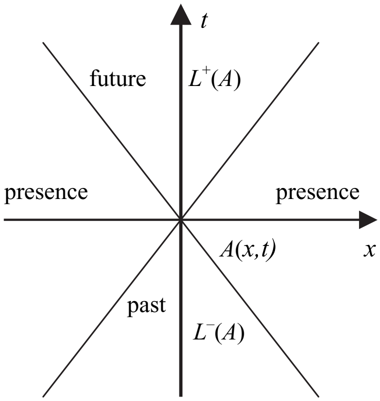


Fig. 9.3. Past, presence, and future of event A

If we apply Lorentz transformations we find that the light cone remains invariant. Hence the future of an event A is transformed into the future, the past into the past and the presence into the presence. The name "presence" for the large spacetime region can now be explained: For every event $G(x_k^G,\ t^G)$ in the presence of $A(x_k^A,\ t^A)$ there exists an inertial system $I'(x_{k'}, t')$ such that the time values $t^{G'}$ and $t^{A'}$ are equal and simultaneous in the usual sense. Simultaneous events are said to be in "space-like" distance. They cannot be connected by causal interaction.

Special relativity cannot say whether there are causal connections between two events E and E' in the Minkowski spacetime. However, if there are causal connections induced by dynamical laws then Special Relativity tells us that the events E and E' have "time-like" distance, which means that E is lying in or on the surface of the backward – or forward light cone of E'.[87]

[87] In spite of this clear situation, in the literature we find a very controversial debate concerning the possibility of backward causation (or retrocausation) in physics. In 1974 Earman (1974, PDT) argued that classical electrodynamics could provide examples for retrocausation. In particular, radiation damping was assumed to lead to pre-acceleration and backward causation. In several papers Grünbaum (1976, MRE) could show, that the "myth of retrocausation" is based on a misinterpretation of the Lorentz–Dirac equation of motion of a charged particle. From a formal point of view the situation is quite clear. The motion of a charged particle is correctly described by the coupled Maxwell–Lorentz equations of particle and field that show no pre-acceleration or backward causation. Only if the field variables are eliminated in terms of particle variables, additional runaway solutions appear which must be eliminated by conveniant initial conditions. For details see the monograph by Rohrlich (1965, CCP), pp. 134–153 and the textbook by Thirring (1979, CMP), Chap. 2.4.

(2) Properties of the relativistic causality relation

From these short remarks we can draw some interesting conclusions for the concept of causality:

1. As in Newtonian spacetime causality is continuous. If $A(x_k^A, t^A)$ is in the past of $B(x_k^B, t^B)$ and $A'(x_k^{A'}, t^{A'})$ is in the future of A but $t^{A'} < t^B$, then A' is also in the causal past of B.
2. As in the Newtonian spacetime the cause event A is always earlier than the effect event B, i.e. $t^A \leq t^B$, in accordance with the chronology condition. However, if we were given two events B and A whose time values t^B and t^A fulfil the inequality $t^A < t^B$, we could in general not infer that A is a possible cause of B. The events A and B could have space-like distance.
3. Since causally connected events have time-like distance, a causality relation is *objective* in the sense that it holds for any observer. Since a cause is lying in the backward light-cone of the effect and since the light cone structure is Lorentz invariant, it follows that a causality relation between two events is preserved under the change of the inertial system. This result is important but surprising only at first glance. For the derivation of the Lorentz transformation we can proceed in two steps. In a first step we derive a transformation which fulfils the requirement of relativity. In this transformation which is not yet fully determined we have still free choice between two options: First, a transformation according to which the temporal order of two events is never invariant and second, the Lorentz transformation which leaves the temporal order of two events with time-like distance invariant. In order to preserve at least the temporal order of causally connected events, Special Relativity makes use of Lorentz transformations.[88] Hence, it should not be a surprise that the causality relation is Lorentz invariant.

(3) Digression: Superluminality

Causally connected events A and B that can be used for the propagation of signals have time-like distance. This means in particular, that we can send signals from A to B at most with the velocity of light $v = c$ in vacuum. The well established Lorentz-invariant theories of classical mechanics and classical electrodynamics confirm this general result. Indeed, particles are moving on spacetime trajectories always with velocities $v < c$. Otherwise their inertial mass would become infinite.

Electromagnetic waves, the most common tool for sending signals propagate with velocities $v \leq c$. In classical physics there is no process known that could propagate with a superluminal velocity $v > c$. Nevertheless, we could speculate what would happen, if we were given a process

[88] cf. Mittelstaedt (1996, KLM), p. 93, 94.

that allows for the transmission of superluminal signals[89],[90] with a velocity $v^S > c$. Elementary calculations show, that two events A and B which are connected by superluminal signals have space-like distance, i.e. $c^2(t_B - t_A)^2 - (x_B - x_A)^2 \leq 0$.

This means that the temporal order of the events A and B is no longer invariant but can be changed by a Lorentz transformation to another inertial system. Hence the order of cause and effect could be inverted simply by changing the reference system. There is only one, though speculative way to resolve this obvious inconsistency. If the superluminal signals with the velocity $v^S > c$ are generally available for the scientific community, similar as light or radar signals, then the spacetime metric could be re-established on the basis of these superluminal signals. In this reformulated spacetime theory that merely replaces c by c^S all inconsistencies would disappear.[91] More details about this way of reasoning can be found in the literature.[92]

(4) Causal topology

The light-cone structure of the Minkowski spacetime M allows for clarifying some more formal properties of the causality relation. For two elements $x, y \in M$ we define a binary relation $C \subseteq M \times M$ and say that for two elements (x, y) the relation C holds, $(x, y) \in C$, when some signal can propagate from x to y. In other words, what happens at x can causally effect what happens at y. It is useful to distinguish two relations of causality: If $C(x, y)$ and $x \neq y$ then we write $C^P(x, y)$ and call as before "C" causal relation and "C^P" proper causal relation $C^P \subseteq C \subseteq M \times M$.

The theory of the causal relation is also called causal topology, The Minkowski spacetime is partially ordered by the causal relation C. This implies the following properties of the proper causal relation $C^P \subseteq C$. The relation C^P is

transitive	$C^P(x, y)$ and $C^P(y, z)$ implies $C^P(x, z)$
asymmetric	$C^P(x, y)$ implies $\neg C^P(y, x)$
irreflexive	$\neg C^P(x, x)$,

where we write $\neg C^P(x, y)$ for $(x, y) \notin C^P$.

In accordance with the previously introduced terminology for $C^P(x, y)$ we also say that x is in the past of y or y is in the future of x. If neither $C^P(x, y)$ nor $C^P(y, x)$ hold then we say that x and y are in the presence of each other.

The formal properties of the proper causal relation C^P are in accordance with the more general arguments of Sect. 9.4.1. The causal relation

[89] Several proposals can be found in the Proceeding of the *Workshop on Superluminal (?) Velocities,* Ann. Phys. (Leipzig), 7 (1998).

[90] Nimtz, G. (2003, SLT).

[91] Mittelstaedt, P. (2000, WSS).

[92] Schelb, U. (1998, STA) and the literature quoted there.

C^P is *irreflexive, asymmetric* and *transitive* and this without any recourse to the philosophical arguments mentioned in 9.4.1. This observation shows again that the causality is not a law of physics but that in some theories a causal relation can be defined by the valid laws of nature. In the following section we will describe explicit examples which don't allow the definition of a causality relation

There is still an important result to be mentioned: The invariance group G of the Minkowski spacetime that is composed of

(i) the orthochronous Lorentz group
(ii) the translation group
(iii) the dilatation group

induces the light cone structure of M. The causality relation C based on this light cone structure defines the causal topology and the causality group G^C. If $f : M \to M$ is a one-to-one mapping then we call f a *causal automorphism* if both f and f^{-1} preserve the partial ordering given by C. In this case we have

$$C(x, y) \Leftrightarrow C(fx, fy) \text{ for all } x, y \in M \ .$$

The causal automorphism form a group, the causality group G^C. If M were two-dimensional (1 space coordinate, 1 time coordinate) then G^C would be much larger than G. The reason is that in the case of two dimensions there are non-linear transformations, not contained in G, which leave the light cone structure invariant. However, if M is the full (3+1) dimensional spacetime, then the additional transformations disappear and the well known relation $G \subseteq G^C$ can be sharpened by the relation $G = G^C$. In other words, "causality implies the Lorentz group". This is the content of the *Zeeman theorem*.[93]

(c) General relativity

(1) *Riemannian spacetime*

If gravitational fields are taken into account then there are no systems of inertia in the sense explained above. Inertial frames of reference can be constructed only locally and momentarily as freely falling reference systems. Since gravitational fields cannot be screened off, there are no force free particles that – as in the Minkowski spacetime – would propagate along straight lines in the sense of Euclidean geometry. The substitute are freely falling point-like particles that are free from non-gravitational forces. They don't propagate along straight spacetime trajectories, but along curved trajectories that can be interpreted as time-like geodesics in a four dimensional pseudo-Riemannian spacetime of signature 2. Hence, in the presence of gravitational fields the Minkowski spacetime must be replaced by the Riemannian spacetime of General Relativity.

The detailed structure of this Riemannian spacetime is determined by the distribution of gravitational masses in the world and by initial

[93] Zeeman, E. (1964, CLG).

and boundary conditions. The law which connects the metric of the Riemannian spacetime with the sources of the gravitational field and the boundary conditions is given by Einstein's field equations. We will not go into details here and refer to Sect. 6.5.6 of this book and the literature quoted there. For the present considerations it is sufficient to assume that we are given a Riemannian spacetime which provides a guiding field[94] that determines the trajectories of massive particles and light rays.

As to the causality relation, in a finite region of the Riemannian spacetime the situation is not so different from a Minkowski spacetime. We find again a light-cone structure, at least locally. To any event point (x_k, t) a light cone is locally defined. This light cone structure is invariant against a large class of spacetime transformations,[95] such that the sequence of past, presence, and future is preserved. Hence different observers with different frames of reference will observe locally the same chronological order of events with time-like distance.

(2) *The large scale structure of spacetime*

(α) The future event horizon

The situation becomes more interesting if we consider not only finite regions of the Riemannian spacetime but global solutions of Einstein's field equations which describe the entire universe in its complete extension in space and time. There are many cosmological models of this kind given by a large variety of Riemannian spacetime manifolds. In the Minkowski spacetime we know, that in a given inertial system an observer O who is located at a fixed space point $P(x_k,t)$ can receive signals from all event points Q of the universe. At any instance t of time the observer can receive signals from all events lying in the backward light cone $L^-(x_k, t)$. Also from all other event points $Q'(x'_k, t')$ outside this light cone signals can be received. In this case the observer O has nothing else to do than to wait – and this means to proceed on its vertical world line (Fig. 9.4). One day, when the observer arrives at an event point $P^*(x_k, t^*)$ the event $Q'(x'_k, t')$will be lying in or on the surface of the backward light cone of the observer O, i.e. in his causal past.

The large scale structure of a pseudo Riemannian spacetime does not generally provide this possibility. In the Minkowski spacetime any future directed time-like geodesic approaches a point i^+, called "*future time-like infinity*" and originates at i^-, the "*past time-like infinity*". In i^+ all possible information about the world can be received by

[94] The expression "guiding field" was introduced by H. Weyl and is further discussed in Chap. 6.

[95] For the restrictions cf. Sect. 6.4.7.2.

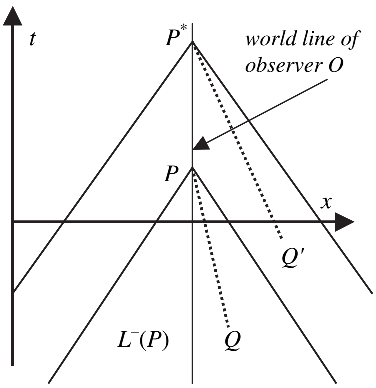


Fig. 9.4. Minkowski space. An observer at a fixed space point receives signals from all events of the universe

realisable signals.[96] The same global structure can be found in many Riemannian spacetime models of the universe.

However, there are also cosmological models, i.e. global solutions of Einstein's field equations that do not show this large-scale structure. For example, in a De Sitter spacetime, which locally has many similarities with our universe, the global structure is very different. For an observer P located at x there are regions of spacetime from which another observer Q located at y could never send signals to P, i.e. the observer P can never be influenced by the observer Q in a causal way. The boundary between events that will at some time be observable for P and those that will never be observable for P is called the "*future event horizon*". Consequently, in a De Sitter spacetime this future event horizon prevents the existence of a "doomsday" at which all information about the universe could be received and registered.

(β) Closed time-like curves

Another feature of cosmological models that could violate the causality structure appears when the global solution considered is not *time orientable*. Locally, there is no problem in a Riemannian spacetime since the chronological order is taken from the Minkowski spacetime. The backward light cone represents the past and the forward light cone the future. Hence, causality should hold locally. However the global question is still open since on a large scale closed time-like curves could exist. Some authors claim that the existence of closed time-like curves leads to paradoxes and must thus be excluded.[97] Indeed, one could imagine travelling round such a curve arriving back before one's departure and preventing oneself from starting.

[96] In a theologically way of speaking the future timelike infinity could also be called "doomsday".

[97] Hawking, Ellis (1973, LSS), p. 198.

Obviously, this argument is conclusive only if we presuppose that the space traveller has "free will" – a premise that cannot be formulated in terms of physics. Moreover, the possibility of closed time-like curves cannot be excluded by general arguments of this kind since there exist explicit solutions of Einstein's field equations with closed time-like curves on a large scale, for instance the Gödel universe. Hence we conclude that causality cannot be required generally on a large scale.

(γ) The first cause

Within the framework of cosmological models, i.e. global solutions of Einstein's field equations that describe an expanding universe, the time-like trajectories considered as chains of causes and effects can be traced back to a point where all time-like trajectories coincide. One could guess that this event may be considered as the "*first cause*" in the sense of the traditional philosophy. However, this event is a singularity that must be excluded from spacetime for mathematical reasons. Moreover, according to some rigorous results[98] this kind of singularity is unavoidable under very general conditions.

However, this result though rigorously valid, does not invalidate seriously the search for a *first cause*. The reason is that in the neighbourhood of the singularity general relativity is no longer the correct theory for the description of spacetime and must be replaced by quantum gravity, a theory that combines quantum mechanics and general relativity. More precisely, if we go back in the history of the universe we finally arrive at the Planck area corresponding to the Planck length $l_{Pl} = 1,62 \times 10^{-22}\,\text{cm}$, the Planck time $t_{PL} = 5,4 \times 10^{-44}\,s$, and the Planck mass$m_{Pl} = 2,18 \times 10^{-5}\,\text{g}$. In this area classical physics looses its validity and must be replaced by quantum physics, in the present case by quantum gravity.

Within the framework of this theory the creation of matter can consistently be described by pair production that is induced by fluctuations of a Riemannian vacuum.[99] At first sight it seems that we could continue our search for a *first cause* and ask for a cause of the vacuum fluctuations mentioned. It must, however, be emphasised that for a single quantum mechanical fluctuation a sufficient reason must not be assumed, not even hypothetically – irrespective of the fact that the statistical distribution of fluctuations is governed by a statistical law.[100] For our present problem this means that any search for a *first cause* of the universe must end here. In other words, *there is no first cause belonging to the evolution of the universe.*

---

[98] Hawking, Ellis (1973, LSS), p. 266.

[99] Vilenkin, A. (1982, CUN) *Creation of Universes from Nothing,* Phys. Rev. Lett. 117B, p. 26.

[100] The general problem of statistical laws that hold strictly though the individual processes are not determined by any law will be discussed in Chap. 12.

9.3.2 Causality in Quantum Physics

(a) Schrödinger dynamics

A quantum mechanical object system S – an atom, a nucleus, or an electron – is described by a complex separable Hilbert space $H(S)$. The states of S are given by the set $T(H)_1^+$ of positive trace class one operators and are either mixed states $W \in T(H)_1^+$ or pure states $P[\varphi]$ that are determined by unit vectors $\varphi \in H.P[\varphi]$ is a one dimensional projection operator that projects another vector state ψ onto φ according to $P[\varphi]\psi = (\varphi, \psi)\varphi$, where (φ, ψ) is the scalar product defined in H of the vectors φ and ψ.[101]

The observables are given in H by the bounded, linear, self-adjoint operators. The simplest operator of this kind is a projection operator $P(M)$ that projects onto a subspace $M \subseteq H$, i.e. on a closed linear manifold contained in H. Observables given by self-adjoint operators assume sharp values, either discrete ones or continuous values. Projection operators have only two values, 0 and 1, and correspond to *properties* which can pertain to an object (value 1) or not (value 0). However, the most general type of observable which is not given by self-adjoint operators allows for unsharp values and represents a very important tool for many applications and fundamental questions.[102] We will not use unsharp observables and properties here, since they do not provide new aspects for our present problem, causality in quantum mechanics.

The simplest situation is given if the physical system S is prepared in a pure state given by a vector ψ. Generally, this state is time dependent and we write $\psi(t)$. The law that describes the temporal development of this "state" $\psi(t)$ is given by the *Schrödinger equation*

$$i\hbar \frac{\partial \psi(t)}{\partial t} = H\psi$$

where H is the *Hamilton operator*, an observable in $H(S)$ that corresponds to Hamilton function in classical mechanics. As in this theory H contains all forces acting on the system. If H does not depend explicitly on t the Schrödinger equation can be integrated by

$$\psi(t) = \exp\left[-\frac{i}{\hbar}H(t - t_0)\right]\psi(t_0)$$

where $\psi(t_0)$ is the initial state at the time $t_0 < t$ and $U(t, t_0) = \exp[-\frac{i}{\hbar}H(t - t_0)]$ the unitary time development operator. Hence we find that the state function ψ at time t_0 strictly determines the state function $\psi(t)$ at a later time $t > t_0$. This will be called the *Schrödinger determinism.*

[101] More details about the quantum mechanical formalism can be found in modern textbooks, e.g. Sakurai (1994, MQM).

[102] cf. Busch et al. (1995, OQP) and the literature quoted there.

At first sight one could get the impression that quantum mechanics is a deterministic theory in the same sense as classical mechanics. Indeed, the causality operator $\exp[X(H)(t-t_0)]$ of classical mechanics seems to be replaced here simply by the unitary Hilbert space operator $\exp[-\frac{i}{\hbar}H(t-t_0)]$ of quantum mechanics. There is, however, a most important difference. The state $\psi(t)$ represents the complete set $\sum_t = \{P^i_\psi(t)\}$ of jointly measurable properties P^i_ψ, expressed by projection operators, that at time t pertain to the system in question either positively or negativly. The set Σ_t of these "objectve" properties is always smaller than the set Σ of all possible properties, i.e. $\Sigma_t \subset \Sigma$. In addition, since the state $\psi(t)$is time dependent, also the set Σ_t changes with time. This means that for a certain property $P^*_\psi\ (t) \in \Sigma_t$ that pertains to S at time t the quantum mechanical causality law could become useless and irrelevant if later at a time t' with $t < t'$ the observable P^* is no longer an objective property of S, i.e. $P^* \notin \Sigma_t$. At subsequent time values t, t', t'' we have different sets $\Sigma_t, \Sigma_{t'}, \Sigma_{t''}$ of properties which are objective with respect to the varying states $\psi(t)$, $\psi(t')$, and $\psi(t'')$ of the system. Since objects are not *completely determined* in the sense of Kant[103] for a particular property P^* there is only a restricted law of causality. In the schematic representation of Fig. 9.5, $P^*_\psi(t_0)$ determines completely $P^*_\psi(t_2)$, but nothing can be said about $P^*_\psi(t_1)$, since the property P^* is non-objective with respect to the state $\psi(t_1)$. Hence the causality law is restricted to a few time frames and thus incomplete.

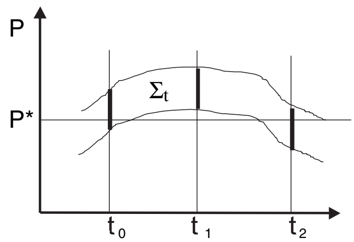


Fig. 9.5. Schematic representation of the temporal variation of the set of objective properties. The causality law is restricted to certain time frames and is thus incomplete.

(b) Statistical causality

These mainly negative statements about the lack of causality in the development of a single quantum system are quite correct. Indeed, if a certain property P^* is not objective with respect to system S in state $\psi_S(t)$, then nothing can be said about whether P^* pertains to S or not. However, irrespective of the correctness of this statement, the property P^* in question

[103] cf. Sect. 9.1.7.

can be “measured”. This means that by a dynamical process the state ψ_S of S is changed (transformed) into a new state ψ_S’ such that P^* is objective with respect to ψ_S’. In other words, the observable assumes either the value 1 or the value 0 which means that P^* pertains to $S(\psi')$ or that P^* does not pertain to $S(\psi')$, respectively.

We will briefly sketch the measurement process here.[104] Let $\psi(S)$ be the state of S before the measurement – the *preparation* – and $\Phi(M)$ the *preparation* of the apparatus M. Furthermore, we consider a discrete observable A with possible values a_i – the eigenvalues – and pure states φ_i^A – the eigenstates – in which the observable A assumes the values a_i. Formally this means $A\varphi_i^A = a_i\varphi_i^A$. Since a measurement is a dynamical process the state transformation can be described by a unitary operator U_M – like the time development operator $U(t, t_0)$ mentioned above. For example, if S is already in an eigenstate φ_i^A and the compound state of $S + M$ reads $\varphi_i^A \otimes \Phi$ then the measurement must lead to the value a_i and the state φ_i^A. Formally this means

$$U_M(\varphi_i^A \otimes \Phi) = \varphi_i^A \otimes \Phi_i$$

where Φ_i is the post-measurement state of the apparatus that indicates the result a_i .

If we apply the measurement operator U_M to the more general situation, when S is in the state ψ, we obtain a new state $\psi(S + M)$ given by

$$U_M(\psi(S) \otimes \Phi(M)) = \psi'(S + M) .$$

This state ψ' which is no longer a product state can, however, be decomposed into a series of product states $\varphi_i^A \otimes \Phi_i$ with coefficients $c_i = (\varphi_i^A, \psi)$ given by the scalar product of eigenstates φ_i^A with the preparation ψ, i.e.

$$\psi'(S + M) = \Sigma_i c_i(\varphi_i^A \otimes \Phi_i) .$$

Hence, by the dynamical process we can achieve that the compound system $S + M$ is in a superposition of product states $\varphi_i^A \otimes \Phi_i$, which means that S is in a mixture of eigenstates φ_i^A but we don’t know in which one.

There is, however an important result which we will briefly mention. Although the state φ_i^A of S and the value a_i of A after the measurement process are completely unknown in the single case, for a large number N of A-measurements of identically prepared systems S^i in states ψ, the relative frequency of the result a_k, say, is given by $|(\varphi_k^A, \psi)|^2$, at least in the limit $N \to \propto$ of an infinite number of trials. Hence, in a measurement process we cannot determine the result a_i in the single case but we can

[104] For more details cf. Mittelstaedt (1998, IQM), Chap. 2.

predict its probability[105] $p(\psi, a_k) = |(\varphi_k^A, \psi)|^2$. This strictly reduced kind of causality was called *statistical causality* by Pauli.[106]

Comparing this result with the preceding section (a) we find that in quantum mechanics there are two serious restrictions of the concept of causality. Either we have the incomplete causality of the Schrödinger dynamics (section a) which is applicable and useful only for a few time-frames, or we consider dynamical measurement processes and find a causality that is complete but only statistically applicable and not relevant in a single case (section b).

As to the general properties of the causality relation we find that the Schrödinger causality (section a) is continuous and the cause is earlier than the effect. By contrast, statistical causality (section b) is not continuous but the cause is earlier than the effect also here.

(c) Superluminality

In spite of the well established results in relativistic mechanics and electrodynamics in quantum mechanics there are several processes that seem to contain among others also superluminal phenomena, i.e. processes which propagate faster than the velocity of light in vacuum. We mention here three candidates for superluminal processes:

- the quantum mechanical tunnelling process,
- the measurement process,
- the EPR correlations.

1. The tunnelling process

   Consider a particle with mass m that is located inside a potential wall as shown in Fig. 9.6.

   More precisely, at time t the wavefunction $\psi(x,t)$ is concentrated in the region between the potential walls, i.e. in region B defined by $a < x < b$. According to the time dependent Schrödinger equation

$$i\hbar \frac{\partial \psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m} \Delta \psi + V(x)\psi$$

   there is a finite probability $p(x',t') = |\psi(x',t')|^2$ to find by a position measurement the particle at a later time $t' > t$ at any point x' in region C. The same argument hlds for region A.

   This well established result of quantum mechanics is confirmed by many experimental results, at first by the decay process of a radioactive nucleus,[107] and then by numerous laboratory experiments. We will not go into details here. The question that is relevant for the causality problem refers to the time that is needed by the particle for the transmission of the potential barrier. Since the velocity of the particle within

[105] For the difficult problem whether in quantum mechanics relative frequencies can be interpreted as probabilities see Sect. 12.2.

[106] cf. Laurikainen, K.V. (1988, BtA), p. 32, 33.

[107] cf. Gamow, G. (1928, ZQA).

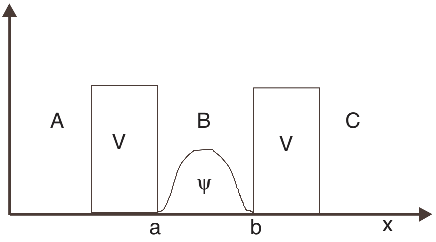


Fig. 9.6. Wavefunction ψ in a potential wall

the tunnel is not restricted by the velocity of light, it seems that we could send superluminal signals using a tunnel barrier. Indeed, we could send a particle from a source S in direction x first tunnelling through a barrier B of width b and finally being registered by the detector D. (Fig. 9.7)

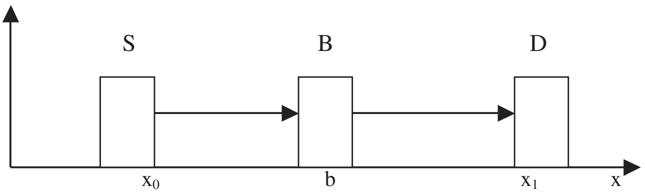


Fig. 9.7. Experimental set-up for the transmission of superluminal signals

The time Δt that is needed by a certain particle to travel from x_0 (source) to x_1 (detector) will be smaller than the time Δt^* which is needed if the barrier is removed. However, it should be emphasised that a complete time dependent description of the transmission process is not yet possible, since a unique time observable is not available in quantum mechanics.[108] There are numerous investigations on this problem but a commonly accepted result could not be achieved.

For superluminal communication the sender (at x_0) and the receiver (at x_1) must agree about a certain time frame $\Delta\tau$. A one-bit message could then consist of sending a particle within the time frame $\Delta\tau$ or not. The receiver can register a particle at x_1 within a slightly delayed time interval $\Delta\tau'$ and thus receive a one-bit signal. Experiments of this kind were realised with massive particles and with photons.[109] There is no doubt that the time δt to cross the barrier is smaller than b/c and perhaps even 0. However, it is also clear that superluminal signals cannot be transmitted in this way. Indeed, a simple quantum

[108] cf. Busch et al. (1995, OQP), p. 77 ff.
[109] cf. Nimtz, G. (2003, OST).

mechanical calculation by means of the Schrödinger equation shows, that a particle emitted at x_0 can either transmit the barrier or it can be reflected at the barrier. There is a well defined transmission probability $p(T)$ and a probability $p(R)$ for reflecting the particle. Clearly we have $p(T) + p(R) = 1$. This means that in case no particle is registered in x_1 within the time interval $\Delta\tau'$, the receiver cannot conclude that the sender emitted no particle. Also the barrier could have reflected the particle. Hence, in this case the receiver does not obtain a reliable signal.

2. The measurement process
Consider again a particle with mass m that at time t_0 is spread over a large region of space, which is described here by the wavefunction $\psi(x, t_0)$. At time $t_1 > t_0$ a position measurement will be performed with the result that S is located at the space-point x_1, say. This is, however, not a stable situation. According to quantum mechanics the time development of the state $\psi(x, t_0)$ leads after a time interval Δt to a state $\psi(x, t + \Delta t)$ with the following properties: Even if the time interval Δt is arbitrary small, for any point x_2 in space there exists a finite probability to register the particle at x_2 by a position measurement. Hence, if S is detected at x_2 it must have moved with a velocity $v(x_1, x_2) = \frac{x_2 - x_1}{\Delta t}$. Since Δt can be made arbitrary small, there is no limitation for the particles speed which can exceed the velocity of light and which could – in principle – become infinite.[110] It should be emphasised that this result holds very general and under very weak assumptions and is not restricted to non-relativistic quantum mechanics.[111]

   This instantaneous spreading of a state can, however, not be used for the transmission of superluminal signals. Let us assume that the sender prepares the particle state by localising it in x_1 at t_1. This particle will now be used as a one-bit signal. If it arrives in x_2 we have yes (1), if not we have no (0). However, in case no particle arrives at x_2 the receiver cannot decide whether the sender emitted no particle or whether the particle is now (at $t = t_1 + \Delta t$) somewhere else. The receiver has only a certain probability to detect the particle at place x_2 but there is no certainty. On the basis of mere probabilities reliable communication is not possible. Hence superluminal signals cannot be transmitted in this way.

3. EPR correlations
The third experimental set-up that is frequently discussed as a means for superluminal communication is the experiment proposed by Einstein,

[110] cf. Schlieder, S. (1971, KRQ).
[111] Hegerfeld G. C. (1980, CLS) and (1989, ISQ).

Podolsky, and Rosen[112] and realised first by Aspect et al.[113] and recently by Gisin et al.[114] Consider two different spin 1/2 systems S_1 and S_2 (e.g. proton and neutron) and assume that the compound system $S = S_1 + S_2$ was prepared in a singlett state $\psi(S)$ with total spin 0. If there is no interaction between the systems S_1 and S_2 the distance between the systems can be made very large. In the experiments of Aspect et al. the distance was 14 m, in the new experiments by Gisin et al. it is almost 10 km.

If the spin observable $\sigma_1(\vec{n})$ in the direction $\vec{n}$ of system S_1 is measured, then the pure state operator $P[\psi(S)]$ of the compound system S is transformed into a mixed state $W(\psi;\, \vec{n})$ that describes the two possible outcomes. There is a strong correlation between the measurement results $\mu\{\sigma_1(\vec{n})$ and $\mu\{\sigma_2(\vec{n})\}$ such that

$$\mu\{\sigma_1(\vec{n}) = \pm 1 \leftrightarrow \mu\{\sigma_2(\vec{n})\} = \mp 1$$

This means that if $\sigma_1(\vec{n})$ was measured with the result $s_1 = +1$, then a measurement of $\sigma_2(\vec{n})$ will lead with certainty to the result $s_2 = -1$.

If the second measurement refers to a spin observable $\sigma_2(\vec{n'})$ with a different direction $\vec{n'}$ then one can no longer predict the result with certainty. In this case quantum mechanics provides only the conditional probability $p(\vec{n};\vec{n'})$ for obtaining the result $\mu\{\sigma_2(\vec{n'})\} = -1$ on system S_2 if $\mu\{\sigma_1(\vec{n})\} = +1$ was measured on system S_1.

Generally, quantum measurements are performed by observers who are equipped with measurement apparatuses. Here we have two observers O_1 and O_2 and apparatuses M_1 and M_2 for measurements of the observables $\sigma_1(\vec{n})$ and $\sigma_2(\vec{n'})$, respectively. We will assume here that the compound system has a large extension and that the subsystems as well as the observers O_1 and O_2 have a macroscopic distance R. In the experiments mentioned the distance R is 14 m and about 10 km. Quantum mechanics does not state that after the first measurement of $\sigma_1(\vec{n})$ with the result $s_1 = +1$ performed by observer O_1 the second observer O_2 has to wait for some time interval Δt before he can obtain the result $s_2 = -1$ of a $\sigma_2(\vec{n})$ measurement with certainty. One could think that this instantaneous action at a distance is not in accordance with the relativistic limitations of velocities between cause and effect. Indeed, since the arguments presented here are based on non-relativistic quantum mechanics, they are not fully convincing. However, the same result can be obtained within the framework of

[112] Einstein et al. (1935, CQD).
[113] Aspect et al. (1982, ETB).
[114] Gisin et al. (2000, OQN).

> Lorentz invariant quantum field theory.[115] Hence from a theoretical point of view there are no doubts in the reliability of our results. Moreover, experimental estimations for the velocity of the causal process between the $\sigma_1(\vec{n})$ measurements and the $\sigma_2(\vec{n})$ measurement led to the result[116] that the cause–effect velocity v_{ce} clearly exceeds the velocity of light and can be estimated by $v_{ce} > {}^2\!/_3 10^7 c$.

There is still an open question. Can the instantaneous correlations in EPR experiments be used for the transmission of superluminal signals? This problem is intensively discussed in the literature since 1970 and led to a result which we will briefly sketch here. Let us presuppose the locality axiom of quantum field theory claiming that two local observables A_1 and A_2 which are measurable by two observers O_1 and O_2 in spacetime regions R_1 and R_2 with space-like distance, commute, i.e. $[A_1, A_2] = 0$. If we apply this axiom to the EPR experiment we find that $\sigma_1(\vec{n})$ and $\sigma_2(\vec{n'})$ must commute. This result has far reaching consequences for the possibility of superluminal signals.

The sender O_1 could try to send a one-bit signal to O_2 by the alternative (measurement of $\sigma_1(\vec{n})$ – no measurement) and the receiver O_2 has to find out whether O_1 performed a measurement or not. By means of a single measurement O_1 cannot send a signal to O_2. If O_1 obtains the result $\mu\{\sigma_1(\vec{n})\} = +1$, say, then O_2 obtains the result $\mu\{\sigma_2(\vec{n}) = -1$. However, this result does not contain any useful information. If O_2 measures $\sigma_2(\vec{n})$ then he will obtain in any case one of the two values ± 1 and it does not matter whether O_1 has performed a measurement or not.

In a next step the sender O_1 could try to send a one-bit signal by performing a *sequence* of $N \gg 1$ $\sigma_1(\vec{n})$ measurements or not. In this case O_1 obtains a sequence of N results $\mu\{\sigma_1\ \vec{n})\} = \pm 1$ with probabilities $p(\pm) = {}^1\!/_2$. However, irrespective of the special results, any $\sigma_1(\vec{n})$ measurement transforms the pure state $\mathrm{P}[\psi]$ of the compound system into the mixed state $\mathrm{W}(\psi, \vec{n})$ mentioned above. The receiver O_2 could try to find out whether or not O_1 has made a series of measurements by measuring the spin observable $\sigma_2(\vec{n'})$ in a different direction $\vec{n'} \neq \vec{n}$ many (N) times. In this way O_2 can determine the expectation value $\sigma_2(\vec{n'})$ with respect to $W(\psi, \vec{n})$. If the expectation value of $\sigma_2(\vec{n'})$ with respect to $P[\psi]$ – without measurement – and to $\mathrm{W}(\psi,\vec{n})$ were different, then O_2 could decide whether O_1 has made a series of $\sigma_1(\vec{n})$ measurements, and in this way receive a one-bit signal.

However, there is an important argument that seems to show that the two expectation values are equal and hence a signal cannot be received in this way. As mentioned above the locality axiom implies that $\sigma_1(\vec{n})$ and $\sigma_2(\vec{n'})$

[115] Schlieder, S. (1971, KRQ).
[116] Gisin N. et al. (2000, OQN), p. 836.

commute. Furthermore, according to a theorem by Lüders[117] it follows that for commuting operators $\sigma_1(\vec{n})$ and $\sigma_2(\vec{n'})$ the expectation values of $\sigma_2(\vec{n'})$ with respect to the states $P[\psi]$ and $W(\psi,\vec{n})$ agree. Hence, O_2 cannot find out whether O_1 has made a series of measurements or not, and consequently O_2 cannot receive signals in this way. – There is, however, one weak link in this chain of arguments. The locality axiom of quantum field theory is an assumption that is usually justified[118] by the argument that it excludes superluminal quantum signals. Obviously, this is a vicious circle and the question whether EPR correlations can be used for superluminal signals cannot be answered in the way described and is still open.

Summarising the results of this section (c) we find that the known quantum mechanical processes that propagate faster than light – tunnelling processes, measurement processes, and non-local EPR correlations – do not allow for superluminal signals. For tunnelling and measuring processes this result is well established. However, in case of EPR correlations the locality axiom must be presupposed, which has no independent justification.

9.4 Do all Laws of Nature Imply Predictability?

Concerning terminology we say that a prediction is a statement about a future event. It is a scientific prediction if it is made in a scientific context with the help of laws and interpreted data about past events and possibly some additional conditions. We say that a law implies predictability if the law together with suitable initial, boundary and possibly other additional conditions, which describe a certain event (state) of some physical system, logically imply predictions which describe some future events (states) of that system.

9.4.1 Arguments Pro

9.4.1.1 If every law of nature describes the time development of state S_1 into state S_2, then this development needs time and cannot proceed faster than with a certain finite velocity. Thus S_2 is always a future state w.r.t. state S_1. But describing a future state means to predict it.

Therefore every law of nature that describes the time development from S_1 into S_2 implies predictability.

9.4.1.2 Every law of nature represents some causal relation (even if not always the same one or to the same extend), as was shown in Sects. 9.2 and 9.3. Now every causal relation satisfies the condition of temporal order such that the effect lies in the future relative to its cause. But this means to predict the effect with the help of the cause and the law.

[117] Lüders, G. (1951, ZMP).
[118] Schlieder, S. (1968, ZML).

Therefore every law of nature implies predictability.

9.4.1.3. According to Chap. 7, all laws are either dynamical laws or statistical laws. Now with the help of dynamical laws we can predict both the behaviour of the whole system and the behaviour of a single part (of a single trajectory); whereas with the help of statistical laws we can predict the behaviour of the whole system (for example towards an equilibrium), but not of a single part.

Therefore all laws of nature imply predictability.

9.4.2 Arguments Contra

9.4.2.1 The motion of dynamical chaos is ruled by dynamical laws. But the motion of dynamical chaos is not predictable. Hence those dynamical laws which underlie and rule the motion of dynamical chaos, do not imply predictability.

Therefore not all laws of nature imply predictability.

9.4.2.2 In quantum mechanics there are processes that proceed with arbitrary velocity and whose final state cannot be predicted. E.g. assume that the position of a particle (electron, proton, etc.) is measured at time t_1 with the result x_1. After an arbitrary small time interval $\Delta t = t_2 - t_1 > 0$ we can find the particle at any place x_2 in space, but the value x_2 cannot be predicted by the laws of quantum mechanics.

Therefore not all laws of nature imply predictability and not all laws of nature require velocities smaller than the velocity c of light in vacuum.

9.4.3 Proposed Answer

Not all laws of nature imply predictability. This is so even if we distinguish different types of predictability according to the differences of dynamical and statistical laws.

9.4.3.1 Different Types of Predictability

The first and strongest kind of predictability is expressed in the quotation of Laplace (Sect. 7.2.1.2). It can be expressed also by the following version which will be called L-predictability:

L-predictability: The state of any closed physical system at any given future instant of time can be predicted with any specified degree of precision, by deducing the prediction from dynamical theories (systems of dynamical laws), in conjunction with initial conditions of which the required degree of precision can always be calculated.[119]

[119] cf. Popper (1982, OUn), p. 36. To avoid misunderstanding we have to mention first that Popper uses the above text to define scientific determinism. Secondly, we skipped the phrase “even from within the system” after “predicted”. This phrase

Observe first that, since dynamical laws are involved, conditions D1 and D2 (Sect. 7.2.3.2) have to be satisfied. In addition, L-predictability contains the following important assumptions:

(i) it refers to any closed physical system,
(ii) it says that any required degree of precision concerning the initial conditions can always be calculated,
(iii) it says that the future state can be predicted with any specified degree of precision.

(i): This condition is too strong, since L-predictability does not refer to thermodynamical systems or systems of radiation or quantum mechanical measurement processes, etc. Therefore (i) has to be restricted, and it seems most suitable to restrict it to systems of classical mechanics.

(ii): Also this condition is too strong in the sense that even a very high degree of precision does not guarantee L-predictability if not in addition condition D4 (Sect. 7.2.3.2) is satisfied.

(iii): The degree of precision depends on the degree of precision with which the initial conditions can be calculated and therefore in turn depends on the satisfaction of D4.

If these restrictions are granted, then L-predictability is possible for classical mechanics and special relativity, as it was shown in detail in Sect. 9.3 above by describing the causal relations in this area of physics. Moreover, if the principle of causality CP1 is satisfied (cf. 9.2.4.1), then also L-predictability is possible. Concerning quantum mechanics, the same holds for Schrödinger dynamics, as it is transparent from Sect. 9.3.2a) above. However, it will be shown below that even in the field of CM there are exceptions for a form of L-predictability which, although it is restricted w.r.t. (i), does not satisfy D4.

Concerning predictability with the help of statistical laws, we have to distinguish three levels of description. According to the distinctions made in 7.2.3.1, the physical system described by statistical laws may be investigated

is especially connected with Popper's interpretation of Laplace, since he interprets Laplace's "intelligence" (usually, but wrongly called "demon") as a human superscientist who cannot "ascertain initial conditions with absolute mathematical precision" (p. 34) and who belongs to the physical world. This weakening of Laplace's intelligence is certainly not extractable from the text and context. But as Earman (1986, PDt, p. 8f.) thinks, Popper uses it to show that also Classical Physics in the understanding of Laplace exhibits some physical systems which are not deterministic. However that may be, we want to stress that the phrase "even from within the system" (i.e. Laplace's intelligence belongs to the physical system) is problematic on independent reasons: As has been pointed out already in Chap. 7, note 14, an intelligence that knows the state in all details (as Laplace says) cannot belong as a part to the physical system. Thirdly, we inserted "dynamical" since it is clear that the laws Laplace had in mind are dynamical laws. Here we use Popper's description only because it contains important assumptions in connection with prediction, as will be discussed below.

only on the macroscopic level. Such is a kind of "phenomenological" approach concerning Classical Thermodynamics. Secondly, the physical system described by statistical laws can be studied at the microlevel with the focus to use the behaviour at the microlevel as the explanatory structure of the behaviour at the macrolevel. In this case the macroscopic magnitudes like temperature or entropy of the system are interpreted as determined by the dynamic behaviour of the microscopic particles. But statistical laws cannot be reduced generally in that sense to dynamical behaviour of single particles as shown in Sects. 7.2.3.4 and 12. below. Therefore on a third level the statistical laws can describe the respective property definitely and objectively w.r.t. to a huge ensemble of particles, although the attribution of this property to the individual particle is objectively undetermined. This happens concerning both, statistical mechanics and quantum statistics.[120] Accordingly we propose two further forms of predictability for statistical mechanics (for example thermodynamics) and for quantum statistics:

S-predictability: The macrostate of a physical system containing a huge number of single objects (particles) can be predicted at any given future instant of time with sufficient degree of precision (in the best case with probability = 1) by deducing the prediction from statistical laws in conjunction with some conservation laws plus initial conditions (about the number of particles, their mean energy, etc.)

Q-predictability: The property attributed to a large number of particles of a quantum system (containing a large number of particles) can be predicted at any given future instant of time with sufficient degree of precision (in the best case with probability = 1) by deducing the prediction from statistical laws in conjunction with QM postulates concerning the measurement process.[121]

9.4.3.2 Chaotic Motion Violates Predictability

We are discussing here the violation of predictability in dynamical systems which show chaotic behaviour in the sense of *dynamical (deterministic) chaos.*[122] However, we shall not go into a discussion of *quantum chaos,*[123] which deals with quantum systems which cannot be described by wavefunctions satisfying Schrödinger's equation. Moreover, we cannot discuss here chaotic phenomena in *large Poincaré systems* (cosmology).[124] The chaotic

[120] For details see Sect. 12.2 below.
[121] For details see Mittelstaedt (1998, IQM), p. 11ff. and 43ff. and Sect. 12.3.2 below.
[122] For different areas of research on dynamical chaos see Berry et al. (eds.) (1987, DCh) and Prigogine (1995, GCh).
[123] For Quantum Chaos cf. Casati, Chirikov (1994, QCh) and Chirikov (1991, TDQ).
[124] cf. Prigogine (1993, TDC) and Fahr (1997, WKS).

motion of both types has properties which differ from those of *dynamical chaos—ff.* The discussion of dynamical chaos will proceed in three steps: First, by discussing an important experiment. Second, by describing properties of chaotic motion. Third, by dealing with the violation of predictability.

(1) The chaotic pendulum
Chaotic behaviour was discovered theoretically (or mathematically) by Hadamard and Poincaré,[125] and experimentally by E. N. Lorenz[126] concerning meteorological and fluid phenomena and only at latest concerning solid mechanical systems, like the pendulum.[127] The experiment with the spherical pendulum shows that even those physical systems which have been understood as ideally satisfying the dynamical laws of classical mechanics, can become chaotic in a strong sense.

The spherical pendulum has a certain type of stability (cf. 7.2.3.2, D4, and 9.4.3.2(1)). Assume we make very small changes in the initial states, say within a neighbourhood distance of ε. Then the distance of the state $h(\varepsilon)$ is proportionally small (no more than a linearly increasing function of time). This kind of stability with respect to small perturbations is called "perturbative stability" which holds in many linear systems. The very important false belief of most scientists until 1970 was that this holds also for the general case.

The important new discovery is now that this simple physical system becomes chaotic if the top end is forced to move back and forth (maximal displacement Δ) with a slightly different period T greater than T_0, provided that Δ is about $1/64$ of l and not more than about a tenth of the energy of motion is dissipated by damping (air resistance, etc.). Miles (1984) showed experimentally that the system is chaotic for values of $T = 1.00234T_0$. It has to be emphasised however that this does not just mean that the system becomes unstable in the sense of simple bifurcation. Instability in the sense of simple bifurcation has been known for a long time. In this case the pendulum weight makes a back and forth oscillation in the same plane and by forcing the upper end this movement begins to be unstable. Such a simple bifurcation, where the plane is not changed, occurs when $T = 0.989T_0$ and slightly above. But for $T = 1.00234T_0$ the pendulum is breaking out of the plane, the number of further bifurcations is arbitrarily increasing, the dependence on initial conditions is completely random such that there is no predictability (or only for very short times).[128]

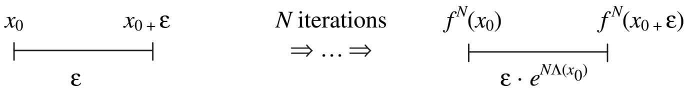
[125] Hadamard (1898, SCO). cf. Ruelle (1991, CCh), Chap. 8. and Popper (1982, OUn), p. 39f. Poincaré (1892, MNM).
[126] Lorenz (1963, DNF).
[127] Miles (1984, RMS). cf. Lighthill (1986, RRF).
[128] In view of this discorvery Lighthill made the acknowledgement cited in Sect. 7.2.3.2, note 34.

(2) Properties of chaotic motion

(a) Sensitive dependence on initial conditions

This is certainly one of the most important characteristics of chaotic behaviour. And it is taken by some as its defining property.[129] But there are phenomena which have that property without being chaotic, as one can see from Maxwell's example (cf. 7.2.3.2, note 37). Therefore this condition (sensitive dependence on initial conditions) cannot be a sufficient condition of chaotic behaviour (motion). But it is certainly an important necessary condition. This important property is measured by the so-called Ljapunov exponent Λ. In fact the Ljapunov exponent measures two things:

(i) It measures the (exponential) separation of adjacent conjugate points (conjugate in respect to the starting point x_0):



This description of stretching of the distance between closely (ε) adjacent points corresponds to a one-dimensional Poincaré map. In real motion the stretching occurs in three dimensions. Then it must be measured by the sum of positive Ljapunov exponents, which is equal to the Kolmogorov entropy (see below).

(ii) The Ljapunov exponent measures also the average loss of information(I_0) about the position of a point in an interval [0, 1] after one iteration. Assume [0, 1] separated into n equal intervals such that x_0 occurs in each of them with probability $1/n$. The answer to the question which interval contains x_0 is then:

$$I_0 = -\sum_{i=1}^{n} \frac{1}{n} ld \frac{1}{n} = ldn \text{ (where } ld \text{ is the logarithm to the base 2)}$$

With decreasing n the information I_0 is decreasing, too, and $I_0 = 0$ for $n = 1$.

Concerning the information presented by a trajectory the Alekseev–Brudno theorem is worth mentioning: The information given with a trajectory of a certain length of time (its algorithmic complexity per time unit) is asymptotically equal to the metric entropy.[130]

(b) Superposition does not hold

Chaotic behaviour in the sense of classical dynamical chaos requires physical systems whose equations are non-linear, i.e. the superposition principle does

[129] "Chaos is thus the prevalence of sensitive dependence on initial conditions, whatever the initial condition is." Ruelle (1990, DCh), p. 42. cf. Farmer (1985, SDP).

[130] cf. Brudno (1983, ECT) and Chirikov (1996, NLH).

not hold since the exponential instability (positive Ljapunov exponent) is non-linear.[131]

It is worth mentioning though that not every chaotic behaviour is non-linear. An example is linear wave chaos in quantum mechanics.[132] Thus non-linearity is a necessary condition for classical dynamical chaos, but for chaotic behaviour in general it is neither necessary, nor sufficient. In general it holds that if the equations of motion are non-linear, then the superposition principle does not hold. However, there are a few exceptions.[133]

(c) Non-periodicity

This property of chaotic motion was already described in Sect. 7.2.3.2 in connection with condition D3.

(d) Bounded motion

Chaotic motion is bounded. This means that the number of degrees of freedom is limited in different ways depending on the kind of chaotic motion. These limitations make the motion – roughly speaking – "oscillatory" in time between stable points (fixed points), whose number multiplies in dependence of certain parameters. One speaks of "folding" with respect to an interval, whereas the exponential separation of adjacent conjugate points (Ljapunov exponent > 0) is called "stretching". Another possibility is that a trajectory becomes attracted to a bounded area of phase space, i.e. to a so-called "strange attractor",[134] within which there is exponential separation of adjacent conjugate points.

Theoretically the boundaries are usually described as Neumann or Dirichlet conditions. In the practical experimental application they may mean for example the size of the system, the number of rolls in a fluid layer of a Bénard experiment, the diffusion coefficients in a chemical experiment etc.

(e) Continuous spectrum

The Fourier spectrum of chaotic motion, which is aperiodic, is continuous and its phase space is continuous (the Poincaré map shows space-filling points, cf. 7.2.3.2, D3). Regular motion on the other hand, which obeys conditions at least D1, D2 and D4, has a discrete spectrum. If we count the number of degrees of freedom (for example these may correspond to the number of

[131] There are famous examples in physics of laws or of the respective physical systems which satisfy the superposition principle. Some of them are the following: acoustic waves, electromagnetic phenomena (time dependent Maxwell equations), optical phenomena, Michelson experiment (independence of the light velocity in respect to moved reference frames). This experiment works only if interference (superposition) is possible. Schrödinger's equation is a linear differential equation. Quantum phenomena are explained with probability amplitudes which can have superpositions. In fact most fundamental equations of QM are linear so far.

[132] cf. Chirikov (1992, LCh).

[133] For example stable solutions of one dimensional non-linear wave equations, so-called solitons.

[134] cf. Ruelle (1980, StA) and the example of the Henon attractor as an explanation of increasing error in Sect. 11.1.3.5(3) below.

rolls of a fluid layer in the Bénard experiment) by the number of Fourier components, then already in an unstable motion at least one such component is continuous. It should be added that quantum chaos violates the condition of the continuous spectrum of the motion and that of continuous phase space. In connection with the uncertainty principle (which allows only a finite size of the elementary cells of phase space) the frequency spectrum of quantum motion is discrete for the motion bounded in phase space.

(f) Change of the variables of the system
A further necessary condition for chaotic behaviour is the permanent change of the important variables of the system. For example in the case of the forced pendulum the change of the amplitude, in the case of a fluid under heat the change of the conductivity of heat in the layers of the fluid. Although these magnitudes remain within a minimum and maximum value they do not recur in the course of time.

(g) Non-integrability
Until Poincaré's discovery to the contrary it was a hidden assumption that all dynamical systems are integrable. And indeed the two body systems like the system sun–earth are integrable in this sense. But if a third body is incorporated (like in the system sun–Jupiter–earth) the system is not anymore integrable. Even less so for the general case of many-body systems. This was the famous prize question of King Oscar II of Sweden of 1885:

> "For an arbitrary system of mass points which attract each other according to Newton's laws, assuming that no two points ever collide, give the coordinates of the individual points for all time as the sum of a uniformly convergent series whose terms are made up of known functions."

The prize was given to Poincaré for his great work "Les Méthodes Nouvelles de la Méchanique Celeste". However, he did not really solve the problem, but gave reasons that such series do not exist, i.e. that contrary to the expectation these series of perturbation theory in fact diverge.[135]

The prize question was partially answered by Kolmogorov in 1954 and solved by his pupil Arnold in 1963. A special case of it was answered by

[135] That a prize should be given for an important mathematical discovery was suggested to the King by the Swedish mathematician Mittag-Leffler. The special prize question was proposed by Weierstrass (the committee consisted of Weierstrass, Hermite and Mittag-Leffler). cf. Moser (1978, SSS). Weierstrass himself was surprised about Poincaré's answer because he arrived (earlier) at the opposite answer. He showed that Poincaré did not in fact prove his result. Today it is known that for very special frequencies such series may in fact converge. cf. ibid. p. 70f. For some of the historical questions concerning that matter cf. the letters of Weierstrass to Sofia Kowalevskaya, Weierstrass (1993, BKW) especially the letter of 15.08.1878, ibid. p. 226ff.

Moser. Hence the name KAM theorem.[136] It gives an answer to the question whether an integrable system (with an arbitrary number of degrees of freedom) survives weak perturbation. The theorem says that the answer is positive and that the invariance with respect to small perturbation or the stability is proportional to the degree of irrationality of the rotation number r (= the ratio of the motion frequency w, to that of the perturbation 2π) of the curve of the trajectory. This has led to a new (weakened) concept of stability which holds for the majority of the orbits; i.e. the majority of solutions (for the respective differential equations) are quasiperiodic.

A criterion of integrability was first given by the Russian mathematician Sofia Kowalevskaya in 1890:[137]

Let $x(t,\, x_0)$ be a function describing the motion of a dynamical system, where x_0 are the initial conditions (the position of the system at $t = 0$). Let the function $x(t,\, x_0)$ have a pole at $t_p = t_1 \pm i\Lambda$ in the complex t plane, where Λ is the Ljapunov exponent.Then the system is integrable if every t_p depends on x_0. Now chaotic behaviour (motion) is non-integrable and therefore its poles do not depend on x_0 and specifically Λ does not depend on x_0.

It should be mentioned, however, that there are weak kinds of chaos or "quasi-chaos", where integrability holds like in Quantum Chaos. A somewhat stronger case is partial integrability (KAM integrability) when an integrable system is exposed to weak perturbation, but is resistant. In this sense integrability (non-integrability) can be used to distinguish levels of disorder in an arrangement beginning with full integrability via KAM integrability to chaos.

Accordingly we may distinguish three levels of increasing disorder and complexity:

(i) Complete integrability: This level of maximal order is characterised by a stable and dynamically predictable motion in terms of individual trajectories (recall conditions D1–D4 of 7.2.3.2).
(ii) KAM integrability: On this level the system is still integrable under sufficiently weak perturbation. Different degrees of disorder on this level correspond to a non-zero Kolmogorov entropy K (see below). In this sense KAM describes degrees of partial integrability. However, the non-integrable part is restricted to an exponentially narrow chaotic behaviour.[138]
(iii) Strong chaos: $K > 0$ and in the limiting case $K \to \infty$. The motion spectrum is purely continuous (cf. (e) above) and the individual trajectory is most complicated like in the example of the chaotic pendulum (cf. (1)

[136] The KAM theorem was proved in: Kolmogorov (1954, CCP), Arnold (1963, SDP) and Moser (1967, CSE).

[137] cf. Chirikov (1991, TDQ), p. 450. For different levels of integrability see Chirikov (1991, PCh) and Prigogine (1993, TDC).

[138] KAM integrability has also an important connection to adiabatic invariance. cf. Chirikov (1987, PCA).

above). Nevertheless the dynamical equations can still be applied and statistical properties of the unstable motion can be proved.[139]

(3) Violation of predictability
As will be clear from the three levels of integrability distinguished above predictability is not violated on the first level; but it is partially violated on the second and completely violated on the third. Since many dynamical systems show weak perturbation and weak chaotic behaviour,[140] the second level of KAM integrability with its different degrees is rather important for the question of predictability. In fact the Kolmogorov entropy (K) can be a measure for predictability. K is a measure of the degree to which a dynamical system is chaotic. For one-dimensional maps, K measures the same as the positive Ljapunov exponent. For higher dimensional systems K is equal to the average sum of all the positive Ljapunov exponents.[141] It was already mentioned that the Ljapunov exponent measures both, the loss of information about the system, and the exponential separation of adjacent conjugate points (recall (2a) above). Moreover, K can be defined by Shannon's measure of information in such a way that K is proportional to the degree of loss of information of the state of the dynamical system in the course of time.[142] Thus it is plain that K is also a measure of the (average) rate for the loss of information of a dynamical system with the evolution of time. In consequence of that, K is also a measure of predictability: it is inversely proportional to the length of time over which the state of a chaotic dynamical system can be predicted.

From the considerations in this Chap. (9.4.3.2) it can be grasped that the following principles are violated if chaotic motion is present:

(i) In case of strong dynamical chaos: Principles of causality CP1, CP4 and L-predictability. On the other hand principles CP2 and CP3 are satisfied. S-predictability and CP4 may partially be satisfied for some statistical properties of the unstable motion.
(ii) In case of weak dynamical chaos corresponding to a certain degree of KAM integrability: CP2 and CP3. The principle of causality CP1, L-predictability and S-predictability may be satisfied at least for some modest positive values of K.

139 cf. Ornstein, Weiss (1991, SPC).
140 For example Laskar (1994, LCS) and others showed that the orbits of the first four planets are partially chaotic; Merkury most (as can be expected because of being closest to the sun), second (unexpectedly) Mars, Venus third, whereas the earth is the most quiet.
141 This was shown by Pesin (1977, CLE).
142 cf. Farmer (1982, IDP).

9.4.4 Answer to the Objections

9.4.4.1 (to 9.4.1.1) It is correct to say that every law of nature which describes the time development of state S_1 into state S_2 implies some kind of predictability. But it is not correct to say that this predictability can be arbitrarily strong or complete. Thus for instance in the case of strong dynamical chaos there is no L-predictability and no L-predictability of a single trajectory except for an extremely short time.

9.4.4.2 (to 9-4.1.2) Although every law of nature represents a causal relation which satisfies temporal order it does not follow from this that the respective causal relation is strong enough, like CP1, to allow L-predictability including predictability of single trajectories. Moreover, the discovery and analysis of dynamical chaos leaves the following question still open: Are there causal relations other than those characterised by the principles CP1 to CP4, which guide the trajectories of chaotic motion in such a way that predictability of them is impossible? Such causal relations may pertain to laws which we have not discovered yet.

9.4.4.3 (to 9.4.1.3) It is correct to say that all laws of nature imply some kind of predictability if by "predictability" it is only meant that some statistical properties of the motion (process) in question can be predicted for a sufficiently short future. This, however, is a very weak statement based on a very weak concept of predictability, although it is then also satisfied by chaotic motion.

10

Laws and Objects

In natural sciences and in particular in physics we usually make use of a dualistic picture for the description of the observed phenomena: There are two clearly distinguished entities, the objects of experience and the laws of nature, where the behaviour of objects is governed by laws of nature. In classical mechanics the spatio-temporal behaviour of mass points is determined by Newton's law of motion. In particular, the trajectories of planets in our solar system are determined by Kepler's laws. The same picture is used in quantum mechanics. The objects, i.e. atoms, electrons, neutrons, etc. are governed by the fundamental law of quantum mechanics, the Schrödinger equation which provides a unitary time development of the quantum mechanical state. The lack of individuality does not invalidate the dualistic schema. The behaviour of a neutron, say, is determined by the Schrödinger equation even if the neutron cannot be individualised. In this case, the law holds for an element of the kind "neutron" but not for a particular one.

In the following sections we will investigate the intricate question whether there are really two well distinguished entities, laws and objects, or whether there is an ontological priority of one of these components with respect to the other one. In physics and philosophy, we find several different answers to this question, some of which will be briefly discussed here.

10.1 Are Objects of Experience Governed by Laws of Nature?

10.1.1 Arguments Pro

The assumption that there are single entities which constitute the reality can be traced back to Aristotle. In *Metaphysics Z* and *H* Aristotle distinguishes its form (*eidos*) and its matter (*hyle*). Generally, we assume that a given object preserves some properties which characterise the object as such, whereas

other properties may or may not pertain to the object without thereby invalidating its persistent identity. A certain stone may change its temperature or change its position in space without thereby loosing the properties that determine this particular stone. Hence, we can distinguish two kinds of properties, the essential properties that determine the object as such and the accidental properties that are varying in time. However, it is not meant, that there is first a well-defined object and in addition essential and accidental properties, which may pertain to it. Instead, according to the Aristotelian ontology the object is constituted by those properties that characterise and determine the object irrespective of the varying accidental properties. Hence, it seems to be correct to identify these essential constituents with the form of the object.[1] In the behaviour of these objects we find regularities which we would call today "laws of nature". In the Aristotelian philosophy the principles[2] (*archai*) which determine the behaviour of the objects, in particular their natural motion, are not reduced to anything else[3] but considered as independent entities.

It is a controversial question whether the form in the Aristotelian ontology refers to a certain kind of objects and hence to a class of things or whether the form refers to an individual object.[4] In the former case an additional principle of individuation is needed for characterising an individual object. In the latter case, individuality of the form means distinguishability of a given object from other objects and re-identifiability at a later time by the form.

The idea that the form determines an individual object was taken up by Duns Scotus and later in particular by Leibniz. For Leibniz an individual object is uniquely characterised by its essential or "internal" properties, which are contained in the "complete concept". Moreover, the "external" properties like position and velocity can be deduced from the complete concept. In addition, the internal properties determine also the history of the object, i.e. its temporal development. This means that in the philosophy of Leibniz the behaviour of individual objects is governed by laws, which follow from the corresponding complete concept. Hence, for Leibniz there is an obvious ontological priority of objects compared to laws (of nature), which appear as derived entities.

Without any explicit recourse to the philosophical tradition in contemporary physics many scientists assume a position which could be considered as a naive realism: There are objects, atoms, molecules, macroscopic bodies, planets, stars, etc. with observable properties, and in addition there are laws that govern the behaviour of the observable properties and in particular the spatio-temporal motion of the objects considered. In other words, objects exist together with properties, even if there are no laws, which govern the behaviour of the variable properties. Hence, objects are considered to be ontologically

[1] Frede, M., Patzig, G. (1988, AMP) p. 44, 45
[2] Wieland (1970, APh), p. 231
[3] Wieland 1.c. p. 64
[4] Frede, Patzig l.c.

prior to the laws of nature, which appear in this picture as independent and contingent structures.

Summarising the philosophical positions of this Sect. 10.1.1. we find that objects of experience are governed by laws of nature.

10.1.2 Arguments Contra

There are serious doubts whether in addition to observed qualities there are some entities, things, or objects, which possess the qualities mentioned as their properties. In his *Treatise of Human Nature* (1739) David Hume emphasised that we never observe objects directly but only qualities and that it is nothing but imagination if we regard the observed qualities as properties of an object. Hence, any scientific cognition begins with the observation of qualities and it seems to be merely a question of interpretation whether in addition to the observed phenomena a fictitious object, "an unknown something"[5] is used for the interpretation and formulation of the experimental results. Consequently, within this scepticism, there is no reason to assume that objects – as fictitious entities – obey at least general laws like the conservation of substance or some causality.

10.1.3 Arguments Contra Hume's Scepticism

The same problem was treated by Kant in his *Critique of Pure Reason* (1787). However, in contrast to Hume, Kant emphasised that "objects of experience" are not arbitrary imaginations but entities that were *constituted* from the observable data by means of some conceptual prescriptions, the categories of substance and causality. Hence, the interpretation of the observed data as properties of an object can only be justified, if an object was constituted as carrier of properties by means of the categories mentioned. Kant formulated necessary conditions, which must be fulfilled by the observational data, if these data are considered as properties of an "object of experience". Accordingly, if we have objective cognition of the reality, i.e. if our observations refer to an element of the exterior reality and not to the observing subject, then the observations in space and time must have been ordered and interpreted according to the categories of substance and causality. In this way, "objects of experience" are constituted and the categories mentioned are necessary preconditions of these objects. It is not claimed by Kant that an interpretation of this kind is always possible. "If each representation were completely foreign to every other, ... no such thing as knowledge would ever arise".[6] If, however, the observations allow for the constitution of objects, then the categories

[5] which view of things ... obliges the imagination to feign an unknown something, or original substance and matter as a principle of union or cohesion among these qualities... Hume (1739, THN), Vol. 1, part IV, Sect. 3.

[6] I. Kant (1787, KRV), A97.

are necessary preconditions of these objects, which fulfil the a priori laws of substance and causality. However, the categories are only preconditions of *kinds* of objects with the same "essential" properties but not of *individual* systems.

In our everyday experience and in the domain of classical physics, which will be discussed in Sect. 10.2.1, to the formal preconditions of experience, the categories of substance and causality, material preconditions can be added which correspond to the material possibilities to measure and to observe properties. These material preconditions of experience specify the formal possibilities for the constitution of objects. In the present case there are no obvious restrictions for measuring all possible predicates P_i jointly on a system which is thus subject to the principle of complete determination:

> "Every thing as regards its possibility is likewise subject to the principle of complete determination according to which if all possible predicates are taken together with their contradictory opposites, then one of each pair of contradictory opposites must belong to it".[7]

A system of this kind or a "thing" possesses each possible "accidental" property P either positive (P) ore negative (not P). In this case, the causality law leads to a strict and complete determination of all properties. In particular, it follows that "things" or objects possess always a well-defined position in space, i.e. they are permanently localised. If in addition impenetrability is assumed, then the permanent localisation can be used for a determination of individual objects by their trajectories in space and time.[8] Later we will find that in the domain of quantum physics Kant's "principle of complete determination" is no longer tenable and must considerably be relaxed.

Summarising the philosophical positions of Sect. 10.1.3. we find that objects exist only in the sense that they are constituted by general laws and that these laws hold a priori for the objects.

10.1.4 Arguments Contra 10.1.3

Kant's transcendental way of reasoning was completely ignored by natural scientists and in particular by physicists in the 19th century. Quite in the spirit of a naive realism mentioned above (10.1.1.) objects were not considered as constituted by categories and other concepts but as really existing material entities. In astrophysics, Kepler's laws govern the motion of real planets, in classical mechanics Newton's law describes the motions and collisions of billiard balls. Even atoms, though not directly visible, were considered by Boltzmann, say, as really existing material entities. Without explicitly mentioning this point, objects obtained again ontological priority with respect to laws of nature.

[7] I. Kant (1787, KRV), B 600.
[8] I. Kant (1787, KRV), A 272.

Obviously, this naive and uncritical position was again exposed to arguments in the spirit of Hume's scepticism. Hence, it is not surprising that the controversy between scepticism and realism was repeated in some sense within the positivism of the 19th century. Ernst Mach[9] argued in a similar way as Hume and denied the existence of atoms since – at this time – atoms could not directly be observed but only inferred from their effects on thermodynamic qualities. Although Boltzmann defended the atomistic viewpoint throughout his career, he has never mentioned Kant's transcendental arguments. Consequently, the Mach–Boltzmann debate could not finally be decided.

10.2 Objects and Laws of Nature in Classical Physics

10.2.1 The Constitution of Objects in Classical Physics

The discussion in the preceding Sect. 10.1 has shown that it is by no means clear whether objects exist and are governed by laws of nature – or whether objects are merely hypothetical entities which allow to interpret the observed qualities as properties of an object. Also Kant's transcendental attempt to combine these two aspects by the idea that objects are constituted by laws was not taken into account in natural sciences.

Hence, in this situation it seems to be correct to investigate in detail the well-known theories of physics and to ask what kind of theoretical entities may be considered as carriers of properties which persist in time. We will discuss this problem at first within the realm of classical physics, since classical physics is often considered as the formal and mathematical representation of our everyday experience which – as such – was subject to philosophical investigations from Aristotle to Kant. In particular, we will study here the theory of classical mechanics since in this theory the basic concepts of classical physics are formulated.

At first sight Mach's critique seems to be correct: Neither in mechanics nor in other classical theories objects appear as genuine entities but only as a means to express the behaviour of measurable quantities. Mechanics deals with the phase space, with observables, canonical transformations, symmetries etc. but not with mass points. Hence, it seems to be interesting to investigate within the framework of classical mechanics the transcendental way of reasoning, which regards objects as entities constituted by categories and laws. Is it possible to constitute in the Kantian way mass points, planets, and stars?

10.2.1.1 Objectivity and Invariance

The Kantian way of reasoning can be made explicit within the framework of classical mechanics in the following way: The goal of physics and in particular

[9] The whole debate is treated fom a historical point of view by Brush (1985, KMH).

of classical mechanics is the cognition of the external reality and not of the observing subject. Accordingly, observations or measuring results should refer to the external reality and not to the observer's subjective impressions. This requirement of objectivity implies that the cognition of the external reality must be independent in some sense of the observer's preconditions. The subjective, observer dependent component of a measuring result is given by the observers spacetime coordinates. Hence, the requirement of objectivity can only be fulfilled if the laws of the external reality have some *invariance* properties. If an observer changes his spacetime coordinates, then the observations should be changed such that they refer to the same but equivalently changed object. In this way, the objectivity of the measuring results can be achieved.[10]

The fundamental laws of classical mechanics are invariant against the transformations of the ten-parameter Galilean group G_{10}. If the observer is "moved" in accordance with a Galilean transformation, the translations in space, say, then the observations that refer to the external object, will transform "covariantly" with respect to this transformation. Since also the observers, represented by measurement instruments are physical objects, they will be subject to the same invariance laws. This implies a symmetry between active and passive transformations: The transformation of the measurement results does not depend on whether the observer is moved according to a Galilean transformation or whether the object is moved according to the inverse transformation.

10.2.1.2 Covariance and Observables

The symmetry between active and passive transformations allows for clarification of the concept of an "observable". Intuitively an observable may be understood as a measurable quantity or a property of an object system S, which belongs to the external reality and which is clearly distinguished from the measuring apparatus. "Properties" correspond to yes–no propositions P_i or to the most simple observables with values 0 and 1. The set $\{P_i\}$ of elementary propositions can be extended by introducing the logical operations $\wedge$, $\vee$, $\neg$, and the relation $\leq$. In this way one arrives at the propositional system of classical mechanics which is given by a Boolean lattice L_C .

One can then define an "observable" in a more formal sense as a relation between numbers on the reading scale of the measurement apparatus and properties of the object system. Hence, an observable may be considered as a mapping Φ

$$\Phi : \text{ß}(\Re) \rightarrow L_C$$

from the Borel sets ß$(\Re)$ of the real line $\Re$ onto the Boolean lattice L_C of propositions. An observable is connected with the group G_{10} of Galilean transformations in a twofold way. Firstly, the properties of the system S are changed by an *active* transformation, when the transformation group acts on the system

[10] H. Weyl (1966, PMN).

and its propositional lattice. Secondly, the reference system of the *observer plus apparatus* is changed by a *passive* transformation, when the transformation group acts on the measurement device M, i.e. on the Borel sets of the reading scale.

Within this conceptual framework the symmetry between active and passive transformations leads to the following important *covariance postulate* (C), which must be fulfilled by an observable: The actively transformed properties of the system S, i.e. the propositions transformed by a representation $S(G)$ of the *Galilean* group coincide with the propositions which are obtained by passively transforming the observers coordinate system by a representation $M(G)$ of the *Galilean* group and hence the reading scale of the apparatus. This means that the diagram in Fig. 10.1 must "commute". The covariance postulate (C) is the abstract formulation of the invariance of classical mechanics with respect to the Galilean group of transformations. It determines those functions which may be considered as "observables" and it shows, how these observables are transformed under a special transformation.

$$\begin{array}{ccc} M & & S \\ \hline \end{array}$$

$$\begin{tikzcd} ß \arrow[r, "\Phi"] \arrow[d, "M(G)"'] & L_C \arrow[d, "S(G)"] \\ ß \arrow[r, "\Phi"] & L_C \end{tikzcd}$$

Fig. 10.1. Covariance diagram of classical mechanics

On the basis of the covariance postulate (C) and the Galilean group, one can now define the fundamental observables p (momentum), q (position) and the observable t (time). In this way the basis quantities (p, q, t) of the state space can be shown to be "observables" in the sense explained, which satisfy the covariance postulate (C). Within the framework of classical mechanics all other observables can be written as functions $F(p, q, t)$ which depend on the coordinates p, q, and t. If an object of classical mechanics is understood as a carrier of properties, then it is obviously sufficient, to require that it is a carrier of the fundamental observables p, q, and t.

10.2.1.3 Classical Objects

One can now define the concept of a classical object S in the following way:

> "A classical object S is given by an algebra L_C, such that a representation of the (passive) Galilean group is defined by automorphism of

the lattice L_C, which admit the observables p, q, t in the sense of the covariance postulate (C) of classical physics."

This means that a classical object is a carrier of the properties $P \in L_C$, not only in one contingent situation K given by an observer, its reference frame and its system of coordinates, but also in all other situations K' which evolve from K by Galilean transformations. The classical object is a carrier of properties which transform covariantly under the transformations of the Galilean group.

One can further specify this concept by considering different classes. For example, elementary systems are given by irreducible representations of the Galilean group. For elementary systems, which correspond to mass points without geometrical structure, there are no *true* but only *projective* representations of the group G_{10}. These representations are characterised by one continuous parameter m that can be interpreted as the "mass" of the object. The next, slightly more general system is a rotating system, with three additional degrees of freedom which correspond to the components of the internal angular momentum.[11]

The result of this Sect. (10.2.1.3) allows for two alternative interpretations. From a realistic point of view, it describes in which way we can recognise an object on the basis of observed properties. From a constructive point of view the mentioned results is merely a definition of what we call an object. For example, Sudarshan et al. writes in this sense:[12] "A free nonrelativistic point particle in classical mechanics *is nothing but* an irreducible canonical realisation of the Galilean group with a positive neutral element."

10.2.1.4 Illustration of the Covariance Postulate: The Objective Reality of the Starry Sky

As a simple example for the covariance principle, which is not trivial, we consider the starry sky. If we are looking to the firmament, we are observing a large number of light spots, which we identify usually with various stars. However, it is not quite clear whether these stars are real entities and not merely our subjective impressions. In order to check the objective reality of the starry sky we follow the way of reasoning mentioned above.

In the four-dimensional Minkowski spacetime M the world line of an observer $B_1(0)$ who is at rest in a given inertial coordinate system is described by a vertical line (Fig. 10.2). The world lines of stars s_1, s_2, s_3, s_4 cross the backward light cone of $B_1(0)$ in event points indicated by black dots. What observer $B_1(0)$ perceives are the light rays coming from these events. The positions of the stars seem to be situated on the celestial sphere that appears to surround B_1. A second observer $B_2(v)$ who is moving at some velocity v

[11] For more details and for more complicated representations of the Galilean group we refer to the literature, e.g. Sudarshan et al. (1974, CDM), pp. 389 ff.

[12] Sudarshan, l.c. p. 391. (emphasis by the authors).

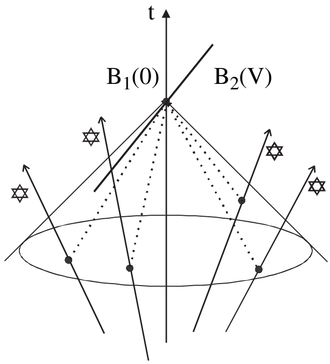


Fig. 10.2. Aberration. Two observers $B_1(0)$ and $B_2(v)$ with relative velocity v perceive different pictures of the starry sky

relative to B_1 and passes by B_1 at the moment when B_1 looks to the sky, will observe the stars s_i to be located in different positions on the celestial sphere. This effect is called "aberration".

The coordinate systems of the observer $B_1(0)$ and $B_2(v)$ are connected by a Lorentz transformation $T_L(v)$. The points $x_i^{(1)}$ where the stars are located on the celestial sphere for observer $B_1(0)$ are different from the positions $x_i^{(2)}$ that are perceived by the second observer $B_2(v)$. The transformations between the positions $x_i^{(1)}$ and $x_i^{(2)}$ are Möbius transformations which map circles into circles and leave angles invariant.[13] For more details about Möbius transformations we refer to the literature.[14] By means of the Lorentz transformations $T_L(v)$ which connect the oordinates of the observers $B_1(0)$ and $B_2(v)$, and the Möbius transformations T_M which connect the star configurations $\sigma_1 = \{x_i^{(1)}\}$ and $\sigma_2 = \{x_i^{(2)}\}$ we can express the condition of objective reality of the starry sky. Observables are mappings $\Phi : \{M_i\} \rightarrow \{\sigma_i\}$ from the subsets $\{M_i\}$ of the Minkowski spacetime to the set $\{\sigma_i\}$ of star configurations. A certain configuration σ_1 which is perceived by $B_1(0)$ has objective reality if the Möbius transformed image $\Phi(M_1)' = \sigma_1' = \sigma_2$ of the subset M_1 agrees with the image $\Phi(M_1')$ of the Lorentz transformed subset M_1', i.e. if the diagram Fig. 10.3 commutes.[15]

---

[13] This exemple is taken from Penrose (1997, LSH), who used it, however, in a completely different context.

[14] H. Weyl (1923, MAR).

[15] Here we used in the covariance diagram the relativistic Lorentz transformation and not the Galilean transformation, since the Lorentz aberration formula is much simple than the corresponding aberration transformation of nonrelativistic celestial mechanics.

$$\begin{tikzcd}
 & M & S & \\
 & B_1(0) \arrow[r, "\Phi"] \arrow[d, "T_L(v)"'] & \sigma_1 \arrow[d, "T_M"] & \\
\textit{Lorentz transformation} \arrow[r, phantom] & B_2(v) \arrow[r, "\Phi"] & \sigma_2 & \textit{Möbius transformation}
\end{tikzcd}$$

Fig. 10.3. The starry sky has objective reality since the covariance diagram of the starry sky commutes

10.2.1.5 Individual Objects

The representations of the Galilean group characterise classes of objects with the same permanent properties. In order to denote an *individual* system one has to find additional properties which distinguish the system S in question from all the other systems $S', S'', \dots$ of the same class. Two questions arise at this point. Firstly, one has to make clear, whether the triple (p, q, t) is a unique denotation of S, i.e. whether there is only one system with these properties. Secondly, if uniqueness is guaranteed, one has to find out in which way the system S defined at time t can be re-identified at some later time $t' > t$. In order to guarantee uniqueness of S one needs an additional dynamical principle which excludes that two systems are at the same time t at the same phase point (p, q). Clearly this postulate is fulfilled if impenetrability in position space is given. This is actually the case in all known situations. However, it does not follow from any dynamical principle. In order to guarantee also the re-identifiability of the system S uniquely defined at time t, at some later time value t', one needs a convenient law which connects the point $(p, q)_t$ in phase space (at time t) with the phase point $(p, q)_{t'}$ (at any other time t'). In classical mechanics a dynamical law of this kind is given by a Hamiltonian $H(p, q)$ and the canonical equations. This means that an individual system S can be re-identified at any other time value $t' \neq t$ by the (p, q) values on its dynamical trajectory $T(S) := (p_t, q_t)$ in phase space. Both requirements for individual objects, the *uniqueness* and the *reidentifyability* are usually guaranteed in classical mechanics. For this reason an individual system S can be named permanently by an arbitrary point (p_t, q_t) on its trajectory $T(S)$.

10.2.1.6 Proposed Answer to Question 10.1 for the Classical World

In classical mechanics there are no primarily given entities that could be regarded as objects. Classical mechanics deals with properties, observables, and

symmetry transformations but not with objects like mass points. Classical objects are given by various representations of the Galilean group, the symmetry group of classical mechanics. Hence, symmetry transformations together with the covariance postulate determine objects as carriers of properties. This way to construct objects may be considered as a realisation of Kant's constitution of objects of experience within the framework of classical mechanics. Mechanical objects, which are constructed in this way fulfil some laws of nature in particular those laws which were used for their constitution, i.e. symmetry laws, covariance requirements, conservation of mass, impenetrability, and Newton's law for trajectories in space and time.

10.3 The Constitution of Objects in Quantum Physics

10.3.1 What is a Quantum System?

The general problems to incorporate objects into our experience which were discussed within the scepticism of David Hume and the positivism of Ernst Mach were sharpened by new difficulties which arise in quantum physics. For this reason, in the first interpretation of quantum mechanics, the Copenhagen interpretation, Niels Bohr assumed a very restrictive position: In this interpretation one considers only measurement results and their mutual relations, but without assuming that the observed predicates can be attributed to an object as its properties. However, Bohr used this "minimal interpretation" not only for philosophical reasons, but because the hypothetical assumption of objects as carriers of properties is sometimes incompatible with quantum mechanics. Indeed, the constitution of objects in quantum mechanics provides problems which are not known from Kant's philosophy and from classical mechanics.

If one tries to extend the Copenhagen interpretation by incorporating objects, then one finds that for quantum systems the laws of substance conservation and causality are no longer generally valid. The reason for this surprising observation is, that quantum systems are not subject to Kant's "principle of complete determination" mentioned above (10.1.3). In quantum mechanics the material preconditions of experience, i.e. the physical laws of measurements, do not allow to determine jointly all possible properties of a given system. In any contingent situation which is described by a state Ψ only a subset P_Ψ of properties can be measured jointly on the system S. The properties $P^i \in P_\Psi$ are mutually commensurable, which means that they can be measured in arbitrary sequence without thereby changing the results of the measurements. The measurement results of these properties (P^i or $\neg P^i$) can be related to the object system just as in classical mechanics. Hence we refer to these properties as the "objective" properties of the system in the state Ψ. However, for any state Ψ there are also *non-objective* properties $P^i \notin P_\Psi$ whose measurement provides a material change of the state Ψ of S. For the non-objective properties Kant's principle of complete determination is violated, since none of a "pair of contradicting opposites" pertains to the object system.

In quantum physics as well as in classical physics for the constitution of objects one has to begin with the requirement of objectivity. The observed predicates should refer to an object as carrier of the respective properties. Again, this requirement leads to the necessary preconditions of any objective experience, the categories of substance and causality. However, in the present case the *material* preconditions of classical experience are not fulfilled, since the systems are not "completely determined". This means that a quantum object system S_Ψ can only be constituted *incompletely* by means of the restricted set of is *objective* properties P_Ψ.

It follows from these arguments that the causality law – in the sense of a dynamical law – holds in quantum mechanics only for the set P_Ψ of objective properties, which are given by the state $\Psi(t)$ at some time value t. The temporal development of this state is determined by the Schrödinger equation in a causal way, i.e. the state $\Psi(t)$ determines the state $\Psi(t')$ at any later time $t' \geq t$. However, since the state Ψ corresponds only to the restricted set P_Ψ of objective properties, at different time values $t', t'', \ldots$ we have different sets $P_{\Psi'}, P_{\Psi''} \ldots$ of objective properties. Hence, it will in general not be possible to establish a causal connection between a property $P^a\ (t)$ at time t and the same property $P^a(t')$ at a later time t'. Consequently, there is only a very limited causality law between the objective properties P_Ψ and $P_{\Psi'}$ at different time values.[16] These restrictions have far reaching consequences for the constitution of objects in quantum mechanics.

10.3.2 Objects in Quantum Mechanics

In principle, the same way of reasoning which allows for the constitution of objects in classical mechanics can also be applied to quantum mechanics. In quantum mechanics as well as in classical mechanics we are interested in the cognition of the external reality and not in the observing subject. This leads again to the requirement of objectivity which means that the fundamental laws of physics are subject to a group of symmetry transformations. Different observers, which are connected by transformations of the invariance group will then be able to describe the same object of the external reality. The invariance group is again given by the Galilean groupG_{10}. The observer corresponds to a macroscopic and classical measuring apparatus, which is associated with a spacetime coordinate system. For this reason a passive Galilean transformation has a meaning, which is quite similar to the classical case. Different observers represented by measurement apparatuses are connected by transformations of the Galilean group and the measuring results will then transform "covariantly" with respect to these transformations.

Similarly as in classical mechanics also in quantum mechanics observables will be characterised by their covariance with respect to the subgroups of the Galilean group. A Galilean covariant sharp observable can then be defined

[16] cf. P. Mittelstaedt (1984, CNI), (1994, OQP); I. Strohmeyer (1995, QTP).

as self-adjoint operator or a projection valued measure Φ on a homogeneous space (equipped with a Borel algebra $\mathcal{B}$) of some subgroup of G_{10}. Observables of this kind allow for sharp measurements of some properties, they are, however, subject to the well-known complementarity restrictions. The sharp properties of a quantum system S at some time value t which correspond to (sharp) yes–no propositions P_i are given by the subspaces of the Hilbert space of the system, or by the corresponding projection operators with eigenvalues 0 and 1. If the set $\{P_i\}$ of propositions is extended by the quantum logical operations $\wedge$, $\vee$, $\neg$ and the implication relation $\leq$, then one arrives at the complete, atomic and orthomodular lattice L_Q of "*quantum logic*". The operations introduced here are defined as intersection and span of two subspaces and as the orthocomplement. (cf. 13.1)

A quantum mechanical observable Φ can then be defined as a relation between pointer values Z on the reading scale of the measurement apparatus M and properties of the object system S. Accordingly, an observable may be considered as a mapping Φ

$$\Phi : \mathcal{B}(\Re) \to L_Q$$

from the Borel sets $\mathcal{B}$ on the real line $\Re$ onto the propositional lattice L_Q of quantum logic, i.e. as a projection valued measure. An observable is then again connected with the invariance group G_{10} in a twofold way. Firstly, the transformation group acts *actively* on the system changing its properties. Secondly, the transformation group acts *passively* on the measuring outcomes which correspond to the Borel sets of $\Re$. The principle of covariance implies again the equivalence of *active* and *passive* transformations[17,18]. Hence, the requirement of objectivity is fulfilled if the image $\Phi(Z')$ of a transformed pointer value Z' agrees with the transformed image $\Phi(Z)'$ of the pointer value Z, i.e. if the diagram in Fig. 10.4 "commutes".

$$\begin{tikzcd}[column sep=huge]
\underline{M} & \underline{S} \\
\mathcal{B} \arrow[r, "\Phi"] \arrow[d, "\mathrm{M(G)}"'] & L_Q \arrow[d, "\mathrm{S(G)}"] \\
\mathcal{B} \arrow[r, "\Phi"] & L_Q
\end{tikzcd}$$

Fig. 10.4. Covariance diagram of quantum mechanics

The difference between the covariance postulates of classical and quantum physics consists in the different propositional systems L_C and L_Q. As in

[17] C. Piron (1976, FQP), p. 93 ff., pp. 77–90.
[18] P. Mittelstaedt (1995, OQM).

classical mechanics the general concept of an observable can again be specified by the fundamental observables of position, momentum and time.

As in the classical case, also quantum objects will be introduced as carriers of the fundamental properties which correspond to the observables q (position), p (momentum), and t (time). Using the covariance postulate we define the concept of a quantum object in the following way: *A quantum object S_Q is given by an algebra L_Q such that a unitary representation of the (passive) Galilean group is defined by the automorphism of L_Q which admit the observables q, p, and t in the sense of the covariance postulate of quantum physics.* This means that a quantum object is a carrier of the properties $P \in L_Q$, but not only in one contingent situation, which is given by the apparatus and its space time coordinates, but also in all situations which can be obtained by Galilean transformations. Hence the quantum object is a carrier of properties $P \in L_Q$, which transform covariantly under Galilean transformations.

However, in spite of the similarities in the method of constitution, there are striking differences between classical objects and quantum objects that come from the different lattices L_C and L_Q, respectively. The propositional system L_C is a complete, atomic orthomodular and distributive lattice. Hence the object S possesses any property $P \in L_C$ either in the affirmative or in the negative sense, i.e. the object S is "completely determined". In contrast to this well known situation a quantum object S possesses at a certain time value t simultaneously only a limited class of commensurable properties given by elements of a Boolean sublattice of L_Q. Hence a quantum system is (at a certain time value t) only a carrier of a class of mutually commensurable properties. One can again specify this concept by considering different classes. Elementary quantum systems are given by irreducible unitary representations of the Galilean group. For elementary objects there are only projective representations which are characterised by one continuous parameter m which can be interpreted as the mass of the quantum object and which characterises a certain class of objects.

10.3.3 Individual Quantum Systems

The characterisation of individual objects in quantum mechanics provides problems which are unknown in our everyday experience and in classical physics. The reasons for the difficulties to individualise quantum objects are that the procedures to determine individual systems which were discussed in the traditional philosophy and in classical physics cannot be applied to quantum objects. The characterisation of individual quantum systems by their essential and permanent properties fails, since the permanent properties define classes of objects (electrons, protons, etc.) which contain more than one element. The characterisation of individual systems by their accidental properties fails since the totality of accidental properties which were needed for the individualisation is not simultaneously available. Since, roughly speaking, only one half of the classical phase space properties pertain simultaneously to

a quantum system and are thus available for the observer, the determination of quantum systems by their accidental properties will never be complete. This means in particular that individual quantum systems cannot be determined by their trajectories in space and time.

10.3.4 Proposed Answer to Question 10.1 for the Quantum World

In quantum mechanics as well as in classical physics there are no primarily given objects. Quantum mechanics is concerned with properties, i.e. subspaces of a Hilbert space, dynamical laws and symmetry transformations. Objects are given by representations of the Galilean group, the symmetry group of quantum mechanics. Hence, symmetry transformations together with the covariance postulate for observables determine objects as carriers of properties. Objects of this kind are governed by laws of nature since they are constituted as entities which are objective carriers of properties and since properties are ruled by laws of nature. However, in contrast to classical mechanics a law does not hold for a particular object, but for a kind of objects, like an electron. Hence, the Schrödinger equation together with its initial conditions holds for an electron, say, but this object can be replaced by any other object of the same kind. The indistinguishability of quantum objects of the same kind corresponds to the permutational invariance of a compound (many body) system, which we discussed in Sect. 5.3.2. For individual objects, there are no laws in quantum physics. (cf. also 12.2)

11

Completeness and Reliability

Assume that we are given a large number of physical laws referring to the well-known fields of physics. Two questions arise:

Firstly, is it possible that the set of physical laws, which are presently known, or a larger set which might be discovered in the not too distant future, is complete. Completeness means in this context that the laws are sufficient in principle to describe any conceivable physical situation, and that no new logically independent law can be found.

Secondly, are the physical laws reliable in the sense that they are true and applicable to realistic physical situations? More precisely, are these laws an immediate image of the processes that we observe in nature or are they merely formal instruments, which can be used for the construction of a correct image of the empirical reality.

We will treat these questions in the following two sections. The question of completeness is concerned with a fundamental structure of laws of nature whereas the problem of reliability refers to the meaning and interpretation of laws of physics.

11.1 Can the Laws of Nature (Physics) be Complete?

11.1.1 Arguments Pro

11.1.1.1 The laws of nature known so far (in the sense of L3, Chap. 1) are always laws of a certain area like those of classical mechanics, of special relativity, of general relativity, or of quantum theory. But inside that area, they are like axioms. From them together with initial conditions one can deduce every truth about that area. But this is just what completeness means: a system of axioms or laws (plus initial conditions) is complete with respect to an area A if and only if all truths about A follow logically from the axioms or laws (plus initial conditions). Therefore, the laws of nature seem to be complete with respect to the area to which they belong.

11.1.1.2 A law of nature (physics) is not complete if there are some phenomena (experimental results) the true description of which cannot be derived from this law (plus initial conditions). However, in such a case it is always possible to make the law complete again by restricting the area in a suitable way. Therefore, the laws of nature (physics) seem to be complete with respect to the area to which they belong.

11.1.2 Arguments Contra

11.1.2.1 Every physical theory which consists of (a system of) laws (in its kernel) presupposes and uses arithmetic. But as Gödel has shown in 1931 arithmetic is incomplete.[1] Therefore, the (system of) laws of a physical theory must be incomplete.

11.1.3 Proposed Answer

The concept of completeness used here is the rudimentary basis idea that a system of laws (axioms) is complete with respect to a given field (domain, research area) of reality if and only if all true sentences about that field are derivable from the laws (axioms). It should be emphasised however that results on completeness or incompleteness obtained in the field of logic and mathematics, cannot automatically be transmitted to physical laws since it is very questionable whether important necessary conditions for these results (like axiomatic construction, closure conditions, definability of the concept of provability within the system etc.) are available for systems of physical laws. Therefore, no more than the above basic idea of completeness will be used subsequently. In accordance with this basic idea the question of the completeness of laws of nature (law of physics) will be discussed in the following way: First some important views in the history (Leibniz, Thomas Aquinas, and Laplace) will be discussed where the case of Laplace is concerned with dynamical laws. Then the question of completeness will be systematically considered with respect to statistical laws, chaotic motion, EPR, and quantum theory.

11.1.3.1 Leibniz: The Laws of Physics Known are not Complete

According to Leibniz there are four great comprehensive areas of scientific research which can be built up according to the principle of rationalism and of *more geometrico* as axiomatic systems: logic and mathematics, metaphysics, physics, jurisprudence (including ethics). He thought that *in principle* these systems of truths are finitely axiomatisable, consistent and complete though for mankind the finite axiomatisability, the consistency, and the completeness are available only for logic, mathematics, and metaphysics but not for the

[1] Gödel (1931, FUS). cf. (1930, SMR).

other two areas since they include contingent truths.[2] The principle for the consistency of axiomatic systems is according to Leibniz: All finitely analytic propositions are (necessarily) true. According to Leibniz a finitely analytic proposition is one that is either an "identical" (i.e. one of the form A is A or $A \wedge B$ is A like "a rational animal is an animal") or one that can be reduced to an identical proposition with the help of a finite number of steps using definitions.[3] Thus, this principle is a soundness principle of the form:

> Every proposition, which is derivable in the axiomatic system, is (necessarily) true.[4]

Leibniz' principle of sufficient reason is a completeness principle. Leibniz gives different formulations of the principle. But he seems to understand them as different versions of the same underlying principle: "that nothing is without a reason, or that every truth has it's a priori proof, drawn from the notion of the terms, although it is not always in our power to make this analysis."[5] "Generally, every true proposition (which is not identical true or true *per se*) can be proved a priori by the help of axioms, or propositions true *per se*, and by the help of definitions or ideas. ... It is certain, therefore, that all truths, even the most contingent, have an a priori proof, or some reason why they are rather than are not. An this is itself what people commonly say, that nothing happens without a cause, or that nothing is without a reason".[6] The principle of sufficient reason can be formulated at least in three versions: (i) nothing happens without a sufficient reason. (ii) Every true proposition has its a priori proof. (iii) Every true proposition is finitely or infinitely analytic. To understand the last version one has to know two things: (1) Leibniz' understanding of analysis: Analysis is a process of logical inference using definitional replacement and determination of predicational containment.[7] (2) Leibniz' view of two kinds of proof processes: In the area of truths of reason the proof–process is terminating after a finite number of steps, i.e. the proposition proved is finitely analytic. On the other hand, in the realm of contingent truths the proof–process is not terminating after a finite number of steps, i.e. the proposition is infinitely analytic. In both cases however, we have a genuine proof–process intended as an inferential procedure of the most rigorous kind, i.e. a proof of a proposition within a deductive system. Our claim has been substantiated in detail (1983, IMS) is that Leibniz' principle of sufficient reason is a completeness principle stating the completeness of a scientific field built up axiomatically (*more geometrico*). This can be seen already from some of the quotes: Every true proposition has an a priori proof (or is provable

[2] For a detailed discussion of this and the following see Weingartner (1983, IMS) part 2. where the respective passages in Leibniz' works are given.
[3] Leibniz (GPh), Vol. 7, p. 195.
[4] For details see Weingartner (1983, IMS), Sect. 2.4.
[5] Leibniz (GPh), Vol. 2, p. 62.
[6] Leibniz (GPh), Vol. 7, p. 300, 301.
[7] cf. Rescher (1967, PhL) p. 23.

from the axioms in the respective deductive system, which represents one of the four fields (logic, mathematics, metaphysics, physics, and jurisprudence with ethics). But although all the four areas are in principle complete for God who knows all the right axioms, the completeness can be effectively shown by men only in those areas where every true proposition is finitely analytic, i.e. can be traced back in a finite number of steps to the axioms. According to Leibniz this is the case in logic and mathematics and metaphysics which deal both only with "true reasons".[8] However in the areas of physics and jurisprudence with ethics, since these areas contain contingent truths, not every true proposition is finitely analytic, some are, but some are only infinitely analytic. In these cases men cannot trace back the respective true proposition to the axioms (since the steps would be infinite) and he cannot find the right axioms (since they are not simple truths of reason but complicated truths of fact). Thus the axioms and laws of physics (and those of jurisprudence with ethics), which are known by men, are not complete.

An interesting related problem for logic is to investigate Leibniz' idea of *ars inveniendi*, i.e. a method to find laws or axioms or explanations if one has a lot of particular truths and lower level hypotheses within a field of research.

11.1.3.2 Thomas Aquinas: The Laws of Nature about the World (Universe) are not Complete

The invariance with respect to displacement in time (and of place) are the oldest and perhaps most important invariance properties of physical laws and laws of nature in general. It seems that the concept of law (of nature) is violated if these invariance conditions would not be satisfied. "The statement that absolute time and position are never essential initial conditions is the first and perhaps the most important theorem of invariance in physics. If it were not for it, it might have been impossible for us to discover laws of nature".[9] The first philosopher, who seems to have been realised this very clearly was Thomas Aquinas. In his quarrel with Bonaventura at the university of Paris he defended the view that the beginning in time of the world (universe) cannot be proved from universal principles (laws) of (about) this world. Because universal principles abstract from *hic* (here) *et nunc* (now): "We hold by faith alone, and it cannot be proved by demonstration, that the world has always existed.[10] . . . The reason is this: the world considered in itself offers no ground

[8] We know today that Leibniz was right with respect to predicate logic of first order but not with respect to mathematics in general. He found an arithmetisation and decision procedure for syllogistics and seemed to be very optimistic to extend such a method further to mathematical proofs. (cf. Weingartner (1983, IMS) ch. 3.1.

[9] Wigner (1967, SRf) p. 4.

[10] In contrast to Thomas Aquinas Bonaventura argued that an infinite past of the universe is *logically* impossible. This argument can be traced back to Johannes Philoponos *"De aeternitate mundi: contra Proclum"*. (1899, DAM). From

for demonstrating that it was once all new. For the principle for demonstrating an object is in its definition. Now the specific nature of each and every object abstracts from the here and now, which is why universals are described as being *everywhere* and *always*. Hence it cannot be demonstrated that man or the heavens or stone did not always exist."[11] The question whether it can be demonstrated that the world has always existed or that it has a beginning in (with) time – answered differently by competing theories of the universe – is a question about the completeness of the laws of nature. Or at least of those laws we know. Thomas Aquinas' standpoint was that the universal laws of nature (about this world) are not complete with respect to all questions (all truths) about this world. It is not just our insufficient knowledge of the laws of nature what he has in mind, but the true laws itself are incomplete according to him with respect to some special questions. Or in more modern terms: the laws of nature are incomplete with respect to some important initial conditions. This problem plays an important role in the Big Bang Theory of the cosmological evolution with respect to (at least) the "first three minutes".[12] However important experiments since 1965 (Penzias and Wilson) revealed the cosmic background radiation, which strongly supports a finite age (and beginning) of the universe. Even if Thomas Aquinas would perhaps say it is not a proof in the sense of demonstration, i.e. a derivation from true or well confirmed laws, it seems still a strong experimental evidence in support to the Big Bang Theory and for a finite age of the universe. That the question is still open is shown by other theories that claim a universe infinite in time and try to be consistent or even to explain the same experimental data.[13]

11.1.3.3 Laplace: The Laws of Nature Together with one State are Complete

The considerations in the last section showed already that "completeness" as it is used in logic is not sufficient when applied to physical laws. Since from laws alone without initial conditions we cannot derive particular truths. Thus the question is rather whether a certain set of physical laws L plus a certain set of initial conditions I could be complete with respect to a certain field F. But how many initial conditions do we need? Laplace thought that we need only one (the minimum) – more accurately: one state of the whole universe at a certain point of time – if the laws are dynamical (in this sense deterministic) laws. Dynamical laws like Newton's laws of motion characterise a physical system by four important conditions. These have been described and discussed as D1–D4 already in Sect. 7.2.3.2. Laplace was mainly concerned

a historical and a systematic point of view the impossibility of an infinite past is discussed by Whitrow (1978, IIP).

[11] Thomas Aquinas (STh)I, 46, 2.

[12] cf. Weinberg (1977, FTM).

[13] cf. for instance Hawking (1988, BHT) who uses imaginary time and Linde (1990, IQC), whose theory avoids this rather implausible assumption.

with conditions D1 and D2 which are usually taken as the defining conditions for determinism or deterministic laws. Lapace's basic idea was someone (a perfect intellect) knowing all the laws (of the universe) could calculate all its states just by knowing one (arbitrary) state.

> "We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instance an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it – an intelligence sufficiently vast to submit these data to analysis – it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes."[14]

This idea can be illustrated by the following example: Assume a film is made of the world, i.e. of the events happening in the whole universe. After the film is developed we cut it into pieces corresponding to single film – pictures. Now we put the single pictures successively in time (in the order of time) into a long card index box like the cards of a library catalogue. Then one special state of the universe at a certain time t corresponds to such a card (film picture) of the catalogue. One can follow one trajectory across the (perpendicular to the) catalogue cards. Interpreted with the help of this illustration Laplace's determinism means that it suffices to know the law(s) of nature and one single catalogue card (film picture) in order to construct all other cards of the catalogue, i.e. to predict and to retrodict all other states of the universe. This means that Laplace's idea is a completeness claim with respect to the laws of nature: the correct laws of nature (known to a perfect intellect) plus one initial state of the whole universe are together complete in the sense that all true descriptions of states of the universe follow from them.

There are some important points however which show that Laplace is only right in a very restricted sense. In general also dynamic (deterministic) laws plus initial conditions are not complete: (1) If the laws in the Laplace's sense, the differential equations of motion were not defined everywhere then also those laws would be incomplete. (2) Laplace's theory is still incomplete with respect to the question of Thomas Aquinas, i.e. it does not imply a finite age of the universe nor does it imply an infinite one in time. In terms of our illustration, Thomas Aquinas would ask: Can you prove that there is a first catalogue card (a first state)? Or can you prove that there is **no** first card (state). And he would point out that this question cannot be decided with the help of Laplace's laws plus available cards (states). (3) If the universe is finite in time then it is possible that some of the solutions of the equations could not be realised because the laws plus one initial condition allow to determine

[14] Laplace (1951, PEP) p. 4.

more states than can happen in the finite time of the existence of the universe. This then is a sense of over-completeness of laws.

11.1.3.4 Statistical Laws Cannot be Complete

After the discovery of statistical laws in thermodynamics and later in other areas there was a general doubt as to an overall interpretation of the world by dynamical-deterministic laws. One of the first philosophers who noticed that a certain imperfection in all mechanical processes allow to enter chance was Charles Sanders Peirce. "But it may be asked whether if there were an element of real chance in the universe it must not occasionally be productive of signal effects such as could not pass unobserved."[15] Moreover, already Thomas Aquinas pointed out that ethical and legal laws – except those very general and formal ones like "the good should be done, the bad should be avoided" – are valid only for most cases and allow exceptions, i.e. are statistical in character.[16]

Concerning statistical laws in physics, it suffices to mention names like Maxwell, Boltzmann, Clausius, Poincaré, Gibbs, and Carnot in order to remind us of very important new discoveries. Processes of heat, friction, diffusion, radiation, electric transport, absorption, dispersion, transmission, relaxation and phenomena like low of water, flow of glacier, floods, avalanches, lightening, growing, aging, propagation, etc. cannot be described by dynamical laws in Newton's sense. Although some essential features of it, the utilisation of the mechanical viewpoint, applied to a huge collection of very small parts of the system, was adopted and lead to a new theory: statistical mechanics. Details and necessary conditions about statistical laws have been discussed in Sect. 7.2.3.3 above. Bearing these conditions in mind, we can give several reasons to show that the statistical laws are incomplete:

(1) Statistical laws are incomplete because they don't describe individual processes.

From what has been said so far one can grasp already one essential reason for the incompleteness of statistical laws. They describe ensembles and the behaviour of ensembles; they don't describe the individual micro processes and microstates. The resulting realised macrostate (a litre of air or a litre of water at a certain temperature) is in fact an ambiguous thing; it consists of a huge number of microstates of which the realised ones change quickly and permanently.

(2) Statistical laws are incomplete because they don't describe correctly the relative frequency of all sequences of events.

---

[15] cf. Peirce (1960, CPC) 6. 47 and the quotations in 7.2.3.1. cf. also Popper (1965, CaC) p. 213.

[16] cf. (STh) I–II, 96,6.

Consider a Mach-Zehnder interferometer (as described in 12.2) with two semitransparent mirrors and two detectors D_1 and D_2. For a very large sequence of identically prepared incoming photons the statistical laws of quantum mechanics predict that the relative frequency of photons that are registered in D_1 is $1/2$. However, since the individual process is not determined at all, it could happen that even in an infinite chain of events all photons are registered in the detector D_1. These non-random sequences have probability zero but they are not excluded by any physical law. Hence, the statistical laws don't describe the relative frequency of these sequences and are thus incomplete.

(3) The incompleteness of statistical laws increases with the complexity of the system.

Consider the following example. Place a little beetle on the first square of a chess board. with borders (so that it cannot leave the chess board). After some time of running around the beetle will again come back to the first square (recurrence). A measure for it will be the probability of recurrence. Place now 10 beetles on each square. The 640 beetles will run around randomly. The probability for recurrence for an individual state of the whole system (say that all beetles are coming back at the same time to their start position) will of course be much lower. Thus the loss of information about a particular state of the system will increase. This can be interpreted as a kind of incompleteness of the statistical hypothesis (law) with respect to the individual (particular) state of the system.

Imagine now a system of 10^{22} molecules. This is Boltzmann's example and his relevant point here: The loss of information (and with it the corresponding incompleteness) about one particular microstate out of $10^{5\times10^{22}}$ when the statistical hypothesis describes the behaviour of the macrostate is in fact enormous. Since non-recurrence and irreversibility are not equivalent notions one may explain the "arrow of time" more modestly by non-recurrence or with Boltzmann by "extremely improbable recurrence".

(4) What is the reason for the incompleteness of the statistical laws?

An old question of the logic of probability has been whether probability and statistical laws have to be interpreted only subjectively (lack of knowledge) or can be interpreted also objectively (representing features of reality).

Is it a lack of our knowledge about physical reality (about nature) which permits us only to use statistical laws for the description of certain phenomena instead of dynamical laws? In this case there could be hidden parameters – unknown to as – which would allow a complete description and explanation (in principle) also for individual states with the help of dynamical laws.

Or is it the indeterminism of nature, of physical reality itself, which forces the laws (and theories) to be incomplete. This question is one of the underlying main questions of quantum mechanics.

11.1.3.5 The New Incompleteness Shown by Chaotic Motion

The new discovery with respect to chaotic motion was that even within an area when dynamical laws are applicable the behaviour of the natural system can change radically such that the motion of a mechanical system may become chaotic. Thus even simple mechanical systems which have been paradigms for strictly obeying dynamical laws (like Newton's law of motion) can turn into chaotic motion and become completely unpredictable just by changing slightly the initial conditions. Since the general characteristics of chaotic motion – mainly in the sense of dynamical chaos[17] – have been discussed in detail in Chap. 9 this section will be concerned only with the aspect of the special kind of incompleteness shown by chaotic motion (in the sense of dynamical chaos). We shall discuss this type of incompleteness in three points: First (1) we shall deal with the sensitive dependence on initial conditions with respect to the loss of information. Second (2) we shall show some important differences between this kind of incompleteness and the one of statistical laws and third (3) we shall discuss increasing error as a negative effect of incompleteness.

(1) Sensitive dependence. Within the last twenty years important properties of chaotic motion (dynamical chaos) have been discovered. One of the most important necessary conditions is the sensitive dependence on initial conditions. Traditionally this aspect is not new, it was known to Aristotle. "The least initial deviation from the truth is multiplied later a thousendfold."[18] A specific example is due to Maxwell: it shows that the principle "like causes produce like effects" is violated if small variations in the initial conditions produce exponentially increasing effects.[19]

    What these examples describe is a strong sensitivity with respect to initial conditions. And this is one of the most important necessary conditions of chaotic motion. (For more details see 9.4.3.2). Observe however that it is not a sufficient condition too and therefore cannot serve as a definition of chaotic motion (in contradistinction to widespread claims): In Maxwell's example the running of the train in a completely different direction is not a chaotic phenomenon, though some aspects of the crash may have chaotic properties.

(2) Differences with respect to the incompleteness in statistical laws.
    Recalling the four points discussed about the incompleteness of statistical laws (cf. 11.1.3.4 (1)–(4)) the main differences concern points (1) and (3): Concerning (1) the incompleteness in chaotic motion of dynamical chaos is not due to not describing individual processes: The opposite is shown by the experiment with the pendulum (cf. 9.4.3.2). This is also clear from the fact that dynamical chaos is guided by dynamical laws which describe,

[17] So-called "quantum chaos" has different properties. cf. Casati and Chirikov (1994, QCh).

[18] Aristotle (Heav) 271b8.

[19] Maxwell (1982, MAM) p. 13. See the full quotation in 7.2.3.2 (D4).

according to the condition D2 (recall 7.2.3.2) not only the physical system as a whole but also all of its parts or individual objects (even if the individual objects may differ in the classical and quantum mechanical situation and Quantum Chaos is not at stake here). The separation of adjacent points and with it the loss of information occurs in the individual case.

Concerning point (3) of statistical laws no complexity is needed for the chaotic behaviour. Although increasing complexity of the chaotic system will usually increase the Kolmogorov entropy, no complexity is needed to begin with. Whereas in the case of statistical laws their incompleteness only begins on a certain degree of complexity since there are no statistical laws describing individual situations, like one special state of the system with a very low probability, if they are not embedded in huge ensembles.

(3) Increasing error
The Hénon attractor[20] is a particular example of a strange attractor (recall 8. 2), which describes the increasing error or better the sensitive dependence on initial small errors. Thus the Hénon attractor can be viewed as one possible interpretation of Aristotle's observation (cf. the quotation in (1) above).

If x_t and x'_t correspond to the initial data x_0 and x'_0 close to each other, the distance $d(x_t, x'_t)$ increases exponentially with t.[21]

$$d(x_t, x'_t) \cong d(x_0, x'_0) \cdot a^t (\text{where}\, a \approx 1,52)$$

Since $a > 1$, a^t increases exponentially with t, i.e. the error $d(x_t, x'_t)$ increases exponentially with time. This means that small initial errors (small errors in the beginning) which are never completely avoidable in the case of experimental data increase exponentially with time.

11.1.3.6 The Incompleteness Discussed in EPR

In their paper of 1935 Einstein, Podolsky and Rosen (EPR)[22] argued that quantum mechanics is incomplete in the sense that it cannot describe every part of the physical reality. This is, however, not correct since the Gedanken-experiment used in the EPR paper – and its various realisations,[23] – does not demonstrate the incompleteness of quantum mechanics. In order to elaborate this argument in more detail we consider a quantum mechanical two-body

[20] cf. Hénon (1976, TDM). That such attractors exist was proved for the Hénon Attractor by Benedicks and Carlson (1991, DHM).

[21] As long as the distance is small. If the distance reaches the order of the attractor it cannot increase anymore.

[22] Einstein et al. (1935, CQD).

[23] The experiments are due to Freedman et al. (1972, ELH), Aspect et al. (1982, ETB), and Kwiat et al. (1995, IBI). Most recent experiments are described by Weihs et al. (1998, VBI).

system, which is composed of, two spin $1/2$ particles (e.g. a proton and a neutron) in a singlet state. This means that the spin of the compound system is 0. If a spin measurement in x-direction, say, of particle 1 leads to the value $s_x^{(1)} = 1/2$, then we can predict with certainty that a spin-x measurement of particle 2 leads to the value $s_x^{(2)} = -1/2$.

A physical theory will be called "complete" if every element of the physical reality has a counterpart in the physical theory. The advocates of the incompleteness theses are using the thought experiment mentioned together with the following requirements.

(i) (Locality) If the two particles are sufficiently separated from each other then a spin measurement of particle 1 will not influence the other particle 2 in any sense.
(ii) (Reality) If without in any way disturbing a system, we can predict with certainty the value of a physical quantity, then there exists an element of the physical reality corresponding to this physical quantity.

These two claims, usually denoted as "locality" (i), and "reality" (ii) are the basis of the EPR incompleteness fallacy.

Let particles 1 and 2 be sufficiently separated such that according to (i) a measurement of particle 1 does not influence particle 2 in any sense. If a spin-x measurement of particle 1 leads to the result $s_x^{(1)} = \pm 1/2$ then we can predict with certainty that a spin-x measurement of particle 2 yields $s_x^{(2)} = \mp 1/2$. The same argument holds for any other direction, in particular for the orthogonal directions y and z. Since the measurement process of particle 1 has no influence on particle 2 the spin values $s_x^{(2)}$, $s_y^{(2)}$, and $s_z^{(2)}$ pertain to particle 2 as real properties in accordance with (ii) and that irrespective of any measurement of particle 1. However, quantum mechanics cannot predict the values of these really existing spin properties of particle 2 and is thus, according to the definition of completeness, incomplete.

The conclusion formulated in the last sentence is not correct. The way of reasoning is correct up to the point where we concluded that the spin values $s_x^{(2)}$, $s_y^{(2)}$ and $s_z^{(2)}$ pertain to particle 2 as real properties. However, this conclusion implies that there is a three-joint probability $p(s_x^{(2)}, s_y^{(2)}, s_z^{(2)})$ which allows to derive inequalities between the marginal two-joint probabilities $p(s_x^{(2)}, s_y^{(2)})$, $p(s_x^{(2)}, s_z^{(2)})$, $p(s_y^{(2)}, s_z^{(2)})$ and the probabilities $p(s_x^{(2)})$, $p(s_y^{(2)})$, and $p(s_z^{(2)})$. Using some simple algebra from these relations we obtain the full set of four Bell inequalities.[24]

Since these Bell inequalities were shown to be violated experimentally in accordance with quantum mechanics we arrive at a contradiction between quantum mechanics and the claims (i) and (ii) mentioned above. Hence it is not possible to infer the incompleteness of quantum mechanics from these requirements. A detailed investigation of the entire problem shows that for a

[24] cf. Mittelstaedt (1998, IQM) p. 99.

consistent description of the EPR Gedankenexperiment the locality condition (i) must be sacrificed. Without this condition it is, however, no longer possible to derive the incompleteness of quantum mechanics in the way mentioned above.

11.1.3.7 The Incompleteness in Quantum Mechanics

Except of the EPR fallacy mentioned above which does not demonstrate an incompleteness of quantum mechanics, there are two other arguments which seem to show that quantum mechanics cannot be complete in an absolute and rigorous sense. These arguments, which are rather irrelevant for all practical purposes, are concerned with the problem of an internal observer and with the self-referentiality of the measuring process. They are loosely connected with Gödel's incompleteness theorem for an important class of formal systems which problem will be briefly discussed at the end of this section.

(a) The internal observer

Within the quantum theory of measurements the object system S and the measuring apparatus M are treated as quantum mechanical systems.[25] Prior to the measuring process the systems S and M will assumed here to be isolated and to be in pure states φ (S) and Φ (M), respectively. The state of the composite system $S + M$ is then given by the tensor product state

$$\Psi(S + M) = \varphi(S) \otimes \Phi(M) \ .$$

A unitary measurement operator U acts on the state Ψ and leads after an appropriate time to a highly correlated state

$$\Psi'(S + M) = U\Psi(S + M) \ .$$

Since Ψ' is an entangled state the reduced states W'_S and W'_M of S and M, respectively, are in general mixed states which provide only a restricted information about the compound system, since they do not contain the correlations between S and M. Moreover, the mixed states do not admit ignorance interpretation and lead, for that reason, to the problem of pointer objectification.[26]

Here we consider a slightly different situation in which the apparatus M is contained in the object system S as a proper subsystem.[27] The system S is then composed of two subsystems M and R, where $R = S\backslash M$ is the residue, and thus $S = M+R$. The apparatus M is used here for measuring the value A_i of an observable A of the object system S. By Z we denote the discrete non-degenerate pointer observable with values Z_i and states Φ_i and assume that a measurement of A leads to a pointer value Z_i which indicates the measurement

[25] Busch et al. (1996, QTM).

[26] For more details cf. Mittelstaedt (1998, IQM) p. 65 ff.

[27] We shall not distinguish here the apparatus and the observer but rather assume that the measuring apparatus plus observer corresponds to system M

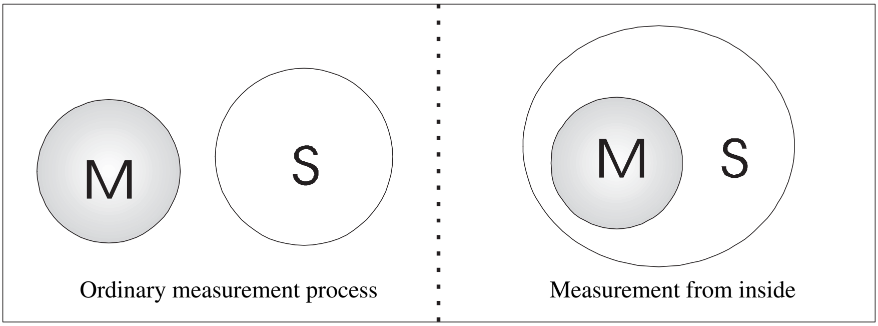


Fig. 11.1. In the ordinary measurement process apparatus M and object system S are separated (*left*), for a measurement from inside the apparatus M is a subset of the object system S (*right*)

result $A_i = f(Z_i)$ where f is the pointer function. A measurement of this kind, called a measurement from inside, is characterised by some peculiarities which will be briefly discussed. (Fig. 11.1)

Since M is properly included in S the object system has more degrees of freedom than the apparatus. This means that there are pairs of different object states $\{W, W'\}$ with $W \neq W'$ that have the same reduced states $W_M = W'_M$ in the subsystem M.[28] Furthermore, if after the measurement of A the pointer is in a state $P[\Phi_\lambda]$[29] and possesses a pointer value Z_λ indicating the object value $A_\lambda = f(Z_\lambda)$ which belongs to the state W^λ then the reduction W_M^λ of this state W^λ to the apparatus M must reproduce the originally measured pointer state $P[\Phi_\lambda]$. This requirement which is fulfilled in the quantum theory of measurement[30] expresses the consistency of the internal measurement process.

However, for measurements from inside the following problem arises. Consider two different states W and W' of the object system S with the same reduced states $W_M = W'_M$ of the apparatus. According to the consistency argument just mentioned, measurements of the two states W and W' (by means of a convenient observable A) lead to the same pointer values and pointer states W_M and W'_M. Hence, although the object states W and W' and the corresponding values of an object observable are different, the internal observer cannot recognise this fact. The pointer states and the pointer values are in both cases the same.

Since the internal observer cannot distinguish all states of the object system, the knowledge that he can obtain about the object system is *incomplete.* Hence quantum mechanics and in particular the quantum mechanical measurement process provides only an incomplete information about the quantum

[28] It is easy to find examples which illustrate this result.
[29] By $P[\Phi\ ]$ we denote the projection operator which projects onto Φ.
[30] cf. Mittelstaedt (1996, IQM), p. 119.

physical reality. However, since this conclusion is relevant only for an internal observer it is rather innocuous. If the object system S is a part of the physical reality, the incomplete information of the internal observer about the system can easily be made complete by an external apparatus plus observer. Hence, the incomplete knowledge of the internal observer is merely a subjective deficiency of information. Only if the object system S is the entire universe the incompleteness of the observer's knowledge cannot be removed. This means that within the framework of quantum cosmology quantum mechanics is incomplete in the described sense. However, it must be emphasised that this deficiency of information is not a particular disadvantage of quantum mechanics, since there are strong arguments[31] that any theory, which is concerned with the *entire* universe leads to similar restrictions of measurability. Hence, one should rather speak of the incompleteness of any cosmology.

(b) Universality and semantic completeness

The assumption of the universal validity of quantum mechanics implies that the theory can be applied not only to any kind of object systems but also to the measuring apparatus plus observer that is used for the experimental justification or refutation of quantum mechanical propositions. Quantum mechanics describes the full measuring process and treats the apparatus as a proper quantum system. The property of a theory to incorporating the means of its own justification, refutation, and interpretation will be called "semantic m-completeness" ("*m*" for measurable), since a theory with this property contains that part of its semantics which is concerned with the confirmation or disconfirmation of statements by *measurements*, i.e. by experimental evidence. Hence "m-completeness" means in this context that the apparatuses which are used for testing the laws of quantum mechanics are not governed by unknown external laws of nature but by the same quantum mechanical laws whose validity they are used to test. Universality implies semantic m-completeness. Meta-theoretical statements, which are concerned with the possibilities and limitations of the apparatus can be reformulated in terms of object theory and proved as propositions of object quantum mechanics. The "interface" between object theory and meta-theory is given by the measurement process.

As an illustration of these remarks we consider an object system S in a pure state φ and an observable A with values A_i. The proposition $A_i(S, \varphi)$: "System $S(\varphi)$ possesses the value A_i of A" can be verified by means of a measuring process. Hence, the meta-proposition "$A_i(S, \varphi)$ is true" which says that $A_i(S, \varphi)$ is verified by a measurement, can be expressed in terms of the object theory by making use of the quantum theory of measurement. If U_A is the unitary measurement operator of the observable A and Φ_i the pointer state which indicates the result A_i, then the translation of the meta-proposition mentioned reads $U_A(\varphi \otimes \Phi) = \varphi \otimes \Phi_i$. A more complicated and also more important example is the problem of joint measurements.

[31] cf. Breuer (1995, IAS) and (1996, SDQ).

The proposition $P := A_i(S, \varphi) \wedge B_k(S, \varphi)$, meaning that "system $S(\varphi)$ possesses the value A_i of A and the value B_k of B" can be verified by a joint measurement process of observables A and B. In general the observables do not commute, i.e. $[A, B] \neq 0$. The meta-proposition "P is true" which says that P is verified by a measurement of the kind mentioned, can be expressed in terms of object quantum theory in the following way. If U_A and U_B are unitary measuring operators for A and B, respectively, e.g.[32]

$$U_A = e^{i\lambda(A \otimes P_A^Z)} \quad \text{and} \quad U_B = e^{i\mu(B \otimes P_B^Z)}$$

which fulfil the calibration condition, then the translation of the meta-proposition mentioned in terms of object language reads

$$U_B(U_A(\varphi \otimes \Phi)) = U_B(\varphi \otimes \Phi_i^A) = \varphi \otimes \Phi_k^B \ .$$

However, this equation can be fulfilled only if the observables A and B commute, i.e. if we have $[A, B] = 0$.[33] Hence, the meta-proposition "P is true" implies that A and B commute.

These arguments apply whenever the apparatus is treated as a proper quantum system. It does not matter here whether object and apparatus are separated (as in section b) or whether the apparatus is included in the object (as in section a). In the ordinary case when the object S and the apparatus M are separated quantum systems, the "probability reproducibility assertion" (PR) can be proved in terms of object theory[34] and the "pointer objectification conjecture" (PO) can be disproved.[35] In case that the apparatus M is contained in the object system S, as discussed in the preceding section (a), the indistinguishability of two states W and W' ($W \neq W'$) with the same reduction to the subsystem M by internal measurements is proved by the object-theoretical statement $W_M = W'_M$. Hence, if quantum mechanics is universally valid in and thus semantically m-complete in the sense mentioned, some meta-theoretical statements can be proved or disproved within the framework of object quantum mechanics.

(c) Relations to Gödel's incompleteness theorem

The situation in the quantum theory of measurement has some similarity with the meta-mathematical problems studied by Gödel.[36] Gödel investigated a

---

[32] Here the parameters μ and λ are proportional to the time interval of the interaction and to the coupling constants that indicate the strength of the interaction. P_A^Z is the observable of the apparatus that is canonically conjugate to the pointer observable Z_A, P_B^Z is defined in a similar way. (For more details cf. Mittelstaedt (1998, IQM) p. 27 ff.)

[33] The special choice of the operators U_A and U_B does not invalidate this result. It can equally be obtained by more complicated unitary operators. (cf. Busch et al. (1995, OQP), Chap. 7).

[34] cf. Chap. 12.3.2.

[35] Mittelstaedt (1996, IQM) p. 122.

[36] Gödel (1931, FUS).

formal system P which fulfils some standard requirements of formalisation and axiomatisation. In particular, the formal system must be rich enough to contain a formalisation (and axiomatisation) of the arithmetic of natural numbers in which primitive recursive functions can be represented; it may appear therefore that the system would contain also all of its own syntax and semantics. That this cannot be the case was shown – among other things – by Gödel's "Incompleteness Theorem".[37] Every formal system P rich enough to contain a formalisation of recursive arithmetic is either consistent or else contains an undecidable (though true) formula. In particular the proof showed that for any formal system P, a sentence S in the language of P equivalent in P to its own P-unprovability (i.e. S is equivalent to nonprovable $\lceil S \rceil$ where $\lceil S \rceil$ is the code for S) cannot be proved in P, provided that P is consistent. Since S is equivalent to non-provable $\lceil S \rceil$, provable $\lceil S \rceil \rightarrow \neg S$ holds in the formal system P. Therefore if: provable $\lceil S \rceil \rightarrow S$ were provable then also: non-provable$\lceil S \rceil$; and consequently S would be provable. Since iteration is accepted as a minimum condition on "provable": provable $S \rightarrow$ provable (provable$\lceil S \rceil$) holds. Thus non-provable $\lceil S \rceil$ and provable $\lceil S \rceil$ would both be provable.[38] From this it follows that every consistent formal system of this kind has limited deductive power and is syntactically and semantically incomplete. For the proof Gödel used a coding system of natural numbers (so called "Gödel numbers") with the help of which the class of axioms and the rules of inference (the relations: x is an immediate consequence of y and z and x is a proof of the formula y) are recursively definable within P. The unexpected result is that a set of the Gödel numbers of all true sentences of a given consistent formal system P belongs to those sets for which no defining expression exists in P. The meta-theoretical results about formal systems in the above sense may be shortly expressed thus: Any formal system containing arithmetic which is such that the proof predicate is recursively definable in it, contains undecidable propositions. Or: For any formal system containing arithmetic: either the proof predicate is not recursively definable in it or it contains undecidable propositions.[39] The up shot is: The price of completeness is inconsistency. An therefore: if consistency has to be granted then one has to put up with incompleteness.

As mentioned above, universality of quantum mechanics implies that metatheoretical propositions can be reformulated in terms of object quantum mechanics making use of the theory of the measuring process. Hence metatheoretical statements play a twofold role. They are elements of the semantics that provides relations between observations and theoretical terms, and they are object-theoretical propositions that follow from the quantum theory of measurement. On account of the analogy with Gödel's work, one could guess

[37] And also by Tarski with respect to questions of definability, in particular also of the notion of truth.

[38] cf. Kreisel (1990, RGC).

[39] Observe that there are of course systems for which both is the case.

that some of these meta-theoretical propositions are not provable within the object quantum theory, even if they can be seen to be true otherwise. If this conjecture were correct, quantum mechanics would be incomplete in a similar sense as arithmetic is incomplete in the sense of Gödel.

However, in spite of this interesting analogy there are important differences between Gödel's investigations and the situation in quantum theory. Since quantum mechanics is not given as a strictly formalised theory it is very hard to see whether the basic requirements of Gödels system P are also fulfilled in quantum mechanics. Hence it is neither clear whether *all* meta-theoretical propositions can be reformulated in terms of object theory, nor is it obvious that some true meta-propositions cannot be proved in the object theory. Gödels important tool that the class of axioms and rules of inference (i.e. the relation of logical consequence) are recursively definable within the formal system by coding the system with natural numbers (Gödel numbering) is not available in quantum mechanics. For this reason we must leave the question open whether in quantum mechanics there is a Gödel-like incompleteness. The incompleteness of quantum mechanics for an internal observer which we discussed in section (a) is of a completely different kind and must not be confused with an incompleteness à la Gödel.

11.1.4 Answer to the Objections

11.1.4.1 (to 11.1.1.1 and 11.1.1.2) Though it is true that all laws of nature known so far are laws of a certain area the restriction and reduction to such an area has a certain threshold, which is connected with two aspects. The first is that the aim of science is to find most general laws with respect to one whole discipline such as the most general laws of physics or of chemistry or of biology. In current research of cosmology or of the theory of evolution, say, the goal is even broader namely to find laws for the development of the whole universe. This is one important reason why the laws restricted to some subarea are viewed as preliminary with the hope of finding more general laws.

The second aspect of the threshold is the necessary condition to preserve severe testability of the laws. To react with restriction of the area of the law after each negative test (in that unrestricted area) leads to immunisation of laws with respect to criticising tests and blocks progress in science. Therefore the aim cannot be completeness of laws via sacrificing generality; and still less the aim can be completeness of laws by buying immunisation and loosing severe testability because this will ultimately sacrifice truth.

11.1.4.2 (to 11.1.2.1) The incompleteness of arithmetic transfers only indirectly and improperly to physics, though physics presupposes and uses (and needs) arithmetic. This is because those arithmetical propositions which are true but are not provable in arithmetic[40] are not interesting or relevant in

[40] A concrete example of a mathematically simple and interesting (arithmetical) proposition which is not decidable in arithmetic was given by Paris and

any reasonable sense for physics. Therefore, the incompleteness of arithmetic does not provide us with special characteristics which are important for the kind of incompleteness of laws of nature (of physics) in the case of dynamical or statistical laws or in the case of special kind of incompleteness in quantum mechanics. The general underlying idea of completeness however, outlined at the beginning of this chapter, is important in all these considerations.

11.2 Are the Laws of Nature Reliable?

11.2.1 What does it Mean that a Law of Nature is Reliable?

A law of nature and in particular a law of physics is called reliable here, if the law holds rigorously, i.e. if under all possible circumstances the law holds and if it provides correct statements about physical facts. Under these conditions, an observer can use the law for predictions and explanations of present and future experimental results. In the sense of this meaning we will raise the question "are the laws of physics reliable?". In particular, we will treat here the following questions:

(1) Do the laws state facts?
(2) Do the laws state single facts?
(3) Do the laws treat isolated situations?
(4) Do the laws describe realistic situations?

11.2.2 Arguments Contra – "How the Laws of Physics Lie"[41]

11.2.2.1 Firstly, we mention the argument that the laws of physics *do not state facts.*[42] By a fact we mean the present state of a physical system, its measurable properties and its position in space and time. Facts are also given by physical processes, which can be observed in a certain region of spacetime. As simple examples we mention the sunrise, a lightning-stroke, or a car accident. Events of this kind are not directly subject to physical laws. The laws of physics describe – without any reference to physical processes – the chronological order of several processes, the mutual influence of processes, and causal connections between events. In particular, laws of nature are only the formal framework, which must be completed by initial conditions, final conditions, causality requirements, numerical values for constants, etc.

In the above-mentioned example of the sunrise the laws of physics which must be applied in this case are the equations of motion of the earth in the

Harrington (1977, MIP). cf. also Kreisel (1980, KGB), p. 175. Later more such propositions have been discovered.

[41] This phrase is adopted from the title of Nancy Cartwright (1983, LPL), *How the Laws of Physics Lie*

[42] Cartwright, l.c. pp. 54–73.

gravitational field of the sun – if we neglect the additional influences of the moon and the other planets. However, the time of the sunrise at a certain place can be determined by these laws only, if in addition the values of the position and of the velocity of the orbital motion around the sun as well as the angle and the angular velocity of the rotation of the earth are given for a certain instant of time. For this reason the laws of physics provide only an abstract scheme which connects possible events but they *do not describe single facts*.[43] In this sense Feynman writes with a slight poetic exaggeration that "there is . . . a rhythm and pattern between the phenomena of nature which is not apparent to the eye but only to the eye of analysis; and it is this rhythm and pattern which we call physical laws. . . ,"[44]

11.2.2.2 Secondly, we could argue that the laws of physics, even if we understand them in the abstract sense mentioned, lie since they treat only *isolated situations* and not real processes. As an example, Feynman formulates the Newtonian law of gravitation as follows:[45] "The law of gravitation is that bodies exert a force between each other which varies inversely as the square of the distance between them, and varies directly as the product of these masses."[46] Even it we accept the very general and abstract character of this law, it does not correctly describe the realistic behaviour of bodies. In fact, if the two interacting bodies possess – in addition to their masses also charges – the motion of the bodies is determined both by the law of gravitation and the Coulomb law between charges. For this reason the law of gravitation formulated by Feynman is not entirely correct. There are two ways out of this dilemma. The first reaction is that Feynman's formulation of the law is not complete and must be completed by a *ceteris paribus* clause as "if there are no other forces than gravitational forces". But then we are again exposed to the argument that the laws of physics describe only isolated, fictitious situations but not realistic ones.

The second reaction is to formulate a new, more general law, which incorporates the Coulomb forces. The two forces, described by vectors $\mathbf{F}_G$ for the gravitational force and $\mathbf{F}_C$ for the Coulomb force could be added making use of the vector addition rule. If this were correct the new, more general law could be formulated by replacing the gravitational force $\mathbf{F}_G$ by the resulting force $\mathbf{F} = \mathbf{F}_G + \mathbf{F}_C$. This way of reasoning is, however, not quite correct. The superposition principle of forces[47] states that whenever two forces $\mathbf{F}_1$ and $\mathbf{F}_2$ are acting on a body, in the equation of motion of this body the resulting force $\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2$ must be used. However, this principle is a new, empirical requirement, which does not follow from the vector addition rule. Without additional assumptions the resulting force reads $\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{f}_{12}(\mathbf{F}_1, \mathbf{F}_2)$

[43] Cartwright, l.c.
[44] Feynman, R. (1967, CPL) p. 13.
[45] Cartwright, l.c. p. 57.
[46] Feynman, l.c. p. 14.
[47] Mittelstaedt, R. (1995, KLM) p. 62.

where $\mathbf{f}_{12}(\mathbf{F}_1, \mathbf{F}_2)$ describes the mutual interaction of the forces $\mathbf{F}_1$ and $\mathbf{F}_2$. If, as in our example $\mathbf{F}_1$ is an electromagnetic force and $\mathbf{F}_2$ is a gravitational force, then $\mathbf{F}_1$ is always influenced by $\mathbf{F}_2$ and vice versa.

Consequently, also the second attempt to show that the laws of physics describe realistic situations fails. Hence it seems that the laws of physics do not describe realistic situations and processes but isolated and artificial cases.

11.2.2.3 According to the arguments in 11.2.2.1 and 11.2.2.2 the laws of physics represent an abstract schema that connects isolated and idealised events but not *realistic situations*. The latter argument can be extended to the objection that the laws of physics – even if they treat isolated cases – are concerned only with artificially simplified cases, which are far from the complexity physical facts. For illustrating this argument we consider the very simple law of free fall in the sense of Galileo which reads in modern terminology $m_I\ddot{x} = -m_G g$. In this law x is the vertical coordinate, m_I the inertial mass and m_G the gravitational mass of the falling body. Since the values of m_I and m_G are proportional with a universal factor we write here the equation of motion in the simple form $m\ddot{x} = -mg$ with $m = m_I = m_G$. The constant $g = 9,81$ cm/sec^2 is the gravitational acceleration on the earth. Integration of this differential equation leads – together with some initial conditions – to the solution $x = -1/2\ gt^2 + v_0 t + x_0$ where v_0 and x_0 are constant initial values. However, taken literally, this law is not correct.

It holds only if the values of the air pressure $p = 0$ (vacuum), the geographic position, the altitude of the experiment, the radius $\rho = 0$ of the falling body (mass point) and other not explicitly known parameters are kept constant. The reason for this condition is that the gravitational force (mg) is not constant but depends on the geographic position, the altitude, and the air pressure. In addition, the law holds only for mass points since additional degrees of freedom of extended bodies would change then simple law of Galileo. Since in the present case for the validity of Galileo's law many parameters which were not mentioned explicitly must kept constant, Cartwright calls it a *ceteris paribus* law.

One could try to incorporate the *ceteris paribus* parameters in a new, more complicated law. This is partially done in Newtons reformulation of Galileo's equation of motion. Hence, the constant g is replaced by kM/R^2, where k is the (Newtonian) gravitational constant, M the gravitational mass, and R the radius of the earth. Hence, the new law reads

$$m\ddot{x} = -\frac{kmM}{R^2}$$

where the number of *ceteris paribus* conditions is strongly but not completely reduced. E.g. since the earth is not an exact sphere, at different places we must use different values of R. One could try to go one step further. In Newton's equation of motion the requirement of special and general relativity are not taken into account. This means that Newton's law does not hold rigorously

but only under the condition that the gravitational field is weak and that the velocities of bodies are very small compared to the velocity of light in vacuum. Hence, Newton's equation of motion is again a *ceteris paribus* law in the above-mentioned sense. The incorporation of the relativity conditions leads finally to Einstein's equation of motion,

$$\frac{d^2x}{d\tau^2} + \Gamma^{\mu}_{\alpha\beta}\frac{dx^{\alpha}}{d\tau}\frac{dx^{\beta}}{d\tau} = 0$$

i.e. to the differential equation of a geodesic line in a four dimensional pseudo-Riemannian spacetime. We will not go into details here and refer to the literature.[48] We stop here since obviously this way of reasoning could be continued up to the formulation of the fee fall within the framework of a "final theory" – if it exists at all. Hence we conclude, that up to now the laws of physics which we find in textbooks and handbooks hold only in the sense of *ceteris paribus* laws.

11.2.2.4 For sake of completeness we should add that there are also *ceteris paribus* conditions of the laws of nature which consist of more formal *hidden assumptions* which we are not aware of. We mention here some examples.

(α) Laws of nature expressed by mathematical formulae presuppose mathematical premises, which are sometimes unknown at the time when the law is formulated. The equation of motion in classical mechanics was considered for a long time as a strictly deterministic equation which allows for a complete prediction of future events. The *ceteris paribus* condition which is not mentioned here consists of restrictive assumptions about the dependence of a solution on the initial condition, since otherwise the motion could become chaotic and thus unpredictible. (For details cf. Chap. 8.)

(β) Concepts are often not sufficiently well defined. To say that a certain equation of motion describes correctly the trajectory of a given mass point presupposes the existence of a symmetry group such that convenient representations of this group allow for the definition of the mass point in question. (For details cf. Chap. 10). For a long time this *group theoretical ceteris paribus condition* was completely unknown and hence a hidden assumption of the kind mentioned.

(γ) There are even logical structures, which are tacitly assumed to hold if a physical theory is applied to some domain of physical reality. If these assumptions were not known when the theory was formulated, then they represent hidden *logical ceteris paribus conditions.* The statement that quantum mechanics, say, is a consistent theory, presupposes – under the conditions of Gödel's theorem – that it is not complete, i.e. that it contains undecidable propositions. (For details cf. Sect. 11.1.3.7c.) Hence, incompleteness is a *logical ceteris paribus condition*, which was completely unknown at the time when quantum mechanics was discovered and a *hidden* assumption of the kind discussed here.

[48] e.g. Rindler (1977, ESR) p. 139.

11.2.3 Answer to the Objections – What does it Mean that a Law of Nature Holds?

In 11.2.2 we mentioned three arguments which show that the laws of physics do not provide a description of real physical events and processes. Instead, the laws of physics are *abstract tools* which are applicable merely to *isolated situations* and *artificially simplified cases.*

To 11.2.2.1: The observation that the laws of physics expressed by equations of motion and field equations *do not describe facts* is completely correct. However, it is a methodological principle of physics for the description of a real process to use two clearly distinguished components: the equation of motion in a very general sense and initial conditions, boundary conditions, causality requirements, etc. In terms of mathematics, this means that a real trajectory, say, is described by a law given by a differential equation and by initial conditions, which correspond to certain values of constants of integration. Thus, the laws of physics together with initial conditions describe facts, although the laws alone don't.[49]

This dualism of laws and boundary conditions has the advantage, that a law can be formulated which "holds" for arbitrary processes and that a special individual process is characterised by additional initial, or boundary values. It is obvious that a law cannot be recognised since any real process is composed of laws and boundary conditions. There is nothing in reality which corresponds to the separation of these two components.

One could go even one step further by incorporating symmetry principles into the description of real processes. The most general component is then the invariance or symmetry principle. The special law is characterised by additional requirements and for the individual process also boundary conditions are needed. However, a real physical process does not show three distinguishable features, a symmetry property, a dynamical law and boundary conditions. These three components exist as separated entities only on the level of our description.

To 11.2.2.2: It is correct to say that the abstract laws of nature together with convenient boundary conditions do not describe arbitrary *real situations* but merely isolated objects and processes. In 11.2.2.2 we found that the two attempts to remove this defect fail. There are, however, good reasons not to change the strategy of physics. First we mention the *context of discovery.* The way in which physical laws can be found at all is to isolate an object from all other bodies, fields and from the environment, and to study the behaviour of this isolated object. (cf. the example of the search for the law of the pendulum in Sect. 5.4.) In this way we arrive at physical laws which are rather simple compared with the laws of compound systems. But even these laws are not always simple in a technical sense. As an example we mention Maxwell's equations in vacuum without gravitation and without incoming radiation and with

[49] For further discussion of the distinction between laws and initial conditions cf. Chap. 8.1.

a finite number of charged mass points as sources of the electromagnetic field. The behaviour of this isolated cloud turns out to be extremely complicated. Even the approximation without free electromagnetic fields leads to Darwin's Lagangian[50], which does not allow for simple solutions. The full relativistic case, which incorporates radiation damping is almost intractable.

Hence, we conclude that in spite of some disadvantages laws of physics must be concerned with isolated and thus unrealistic cases since otherwise they could not be discovered. The fundamental laws of physics are *ceteris paribus* laws since the incorporation of the tacitly assumed conditions would lead to very complicated laws which could neither be discovered nor applied on account of their mathematical intractability.

To 11.2.2.3: Obviously, it is a correct observation that the fundamental laws of physics refer not only to *isolated situations* but also to oversimplified cases. They are *ceteris paribus* laws since influences from various sources are tacitly excluded. The attempt to incorporate the ceteris paribus conditions and to obtain in this way laws for more complex situations leads to more and more complicated laws of physics. This was illustrated for the equation of motion in a gravitational field.

The formulation of more complicated laws for more complex situations is not only complicated but also *not desirable*. We mention here three reasons. Firstly, the incorporation of *ceteris paribus* conditions leads step by step to more complicated laws. It is hard to say whether this sequence ever ends. There are explicitly assumed *ceteris paribus* conditions, which can in principle be eliminated. There are, however also conditions which come from "higher order" theories and which are less known. In addition, we don't know, whether the hierarchy of theories ends after a final number of steps in a "final theory"[51] or in a "theory of everything". Only in this case we could eliminate – at least in principle – the totality of *ceteris paribus* conditions. Hence, this way of reformulating the laws of physics is not very promising.

The second argument refers to the "context of justification". One task of physics is the formulation of laws in terms of mathematics. Another task is the justification of these laws by experimental evidence. The experimental justification is practically possible only for simple laws which hold of course only under various *ceteris paribus* conditions. For example, the quantum mechanical Schrödinger equation could in principle be tested at a complex molecule in an electromagnetic and a gravitational field. From a practical point of view it is much easier to test this equation at a one body problem with Coulomb forces only without gravitation and electromagnetic waves. This is the well-known hydrogen atom problem which can be solved exactly and easily be compared with experimental results. Hence, also within the context of justification simple though somewhat artificial laws have many advantages.

[50] Landau, L.D. Lifschitz, E.M. (1963, ThP) p. 187.
[51] cf. Weinberg (1993, DFT)

There is a third, more sophisticated argument in favour of simple physical laws, which hold for artificial and oversimplified cases. The totality of laws of physics must not be considered as a collection of independent rules. Instead, these laws constitute multiple connected networks of laws, which are usually called *theories*. As an example for a physical theory we mention Maxwell's theory of electrodynamics. In this theory from Maxwell's equations and some other requirements all other laws of electricity and magnetism, e.g. the laws of Faraday, Biot and Savart, and Oersted, can be derived. Theories of this kind are constituted by simple laws which refer to simple, sometimes oversimplified experimental situations. Only the interplay of these multiple connected laws will apply to realistic and complex situations.

In addition, in contemporary physics an even more hypothetical argument is often used. Theories are not considered as logically independent structures but as complex systems of laws, which are on their part connected in a manifold way. Some physicists assume that there is a hierarchy of theories such that from a "theory of everything"[52] all other theories can be derived. We will not discuss here the question whether it is the goal of physics to find a "final theory"[53] and whether this goal can ever be achieved. From the preceding discussion it is, however quite clear that a hierarchy of theories and a possible final theory will only be concerned with abstract and simple laws which do not describe directly our immediate physical experience.

11.2.4 Summary

It is a correct observation that the majority of laws of physics are formulated incompletely without mentioning all constraints and conditions, which are tacitly assumed to be fulfilled. Hence, the physical laws which we find in our textbooks and handbooks are not literally applicable to real physical situations. In this and only in this sense one could say that "the laws of physics lie"[54]. This statement can be further illustrated by comparing physical laws and mathematical laws. In the correct formulation of a mathematical theorem all premises must be formulated explicitly. In this way exceptions can be fully excluded and the theorem holds generally. A formulation of this kind is always possible since the premises of the theorem in question are known. In physics the situation is quite different. Firstly, there are many preconditions, which are tacitly assumed since otherwise the laws would become intractable. Secondly, many preconditions of a physical law cannot be mentioned explicitly since they are not, or more precisely not yet known. In the example of the

[52] The denotation "theory of everything" is ambiguous and somewhat misleading. It is not meant here that a theory of everything describes the totality of observable phenomena but only the fundamental structures which allow to derive the structure and the behaviour of every complex observable phenomenon. cf. also Barrow (1991, TOE). T'Hooft G. (1995, PVW) and Breuer (1997, QMG).

[53] cf. Weinberg (1993, DFT).

[54] Cartwright (1983, LPL).

equation of motion in a gravitational field we found, that the number of conditions increases if we proceed step by step to more general theories. Starting from Galileo's formula for the free fall, we proceeded to Newton's theory and further to Einstein's theory. The next step would be the formulation of this problem within the framework of quantum gravity, etc. This means that the list of preconditions could be completed only, if we were in the possession of a *theory of everything,* i.e. if physics were already in its final stage.[55] This is, however not the case. Also the mathematical and logical hidden assumptions mentioned above could not be taken account of in the original formulation of physical laws, since they were not yet known when the laws were formulated. One could speculate that even today there are formal hidden assumptions which are presupposed in some laws of nature and which we do not yet know.

In spite of difficulties which might arrive if we are dealing with laws that are not directly applicable to realistic situations, there are good reasons to maintain very simple and incomplete physical laws. Laws of this kind have many advantages and they are almost unavoidable

- in the *context of discovery,* since otherwise the laws could not be found. For example, for the discovery of quantum mechanics it was essential that there are a few exact and simple solutions of the Schrödinger equation (hydrogen atom and harmonic oscillator) which could be compared directly with simple experimental situations.
- in the *context of justification,* since experimental tests can only be performed on isolated objects in strongly simplified situations. For this reason most interesting features of quantum mechanics as non-locality and entanglement were tested only in recent years since the experimental technique for investigating isolated, individual quantum objects did not exist before.
- in the *context of explanation*, since only for simple laws we can hope to explain one law by another one. Generally, we want to know which laws are explanatory relevant for which others. In particular we want to explain complex phenomena as the result of the interplay of simple laws. This is, however possible at most for laws which hold for simple and isolated situations. Even for these cases, the problem could become very complicated. In the example of an interplay of gravitational forces and Coulomb forces we found that the vector addition law does not hold due to non-linear terms which arise in this case.

In addition, simple, idealised laws can be used for the formation of theories which can then be connected by intertheoretical relations.[56] In this way a network of theories can be constructed which could serve as a guideline for future research work.

As to the complexity of realistic situations and processes it must be emphasised quite generally, that the laws of physics are not conceived and must

[55] In contrast to S. Hawkings prediction of 1980, (1980, ETP) that "*the goal of physics might be achieved in the not too distant future*".

[56] cf. Scheibe (1997, RPT), (1999, RPT).

not be understood as mathematical simulations of these complex situations. Instead, they must rather be considered as various tools, which can – together with convenient boundary conditions – jointly be used for the description of realistic situations. In other words, the laws and theories of physics are not images of the real world but rather tools for constructing correct models of the world. This means, however, that the physical laws that we find in textbooks, monographs, and handbooks are only the first step. In addition to the formal laws, we must learn to apply them to the real world. This is a difficult task, in particular since there are no general rules for considering or neglecting various *ceteris paribus* conditions of the physical laws which we find in the textbooks. Moreover, the real processes are not pure cases but must be described by a joint application of several laws. Hence, there are many pitfalls in interpreting and understanding the laws of physics.

Part III

Why are Laws of Nature Valid?

12

Statistical Laws

12.1 General Preliminaries

12.1.1 Regularity and Necessity

In our experience we observe regularities. Compared with the period of a pendulum the rotation of the earth is periodic and defines the duration of a day. The orbital motion of the earth around the sun is again periodic and defines the duration of a year. If there are no gaps in these regularities we assume that there is a law behind the observed phenomena that governs the motion of the earth and the other planets in our solar system. The first attempt to formulate laws of the planetary system in terms of mathematics was made by Kepler who was convinced that the three laws formulated by him already elucidate the intrinsic harmony of the universe. This conclusion was, however, somewhat too hasty. A law of physics should not describe a particular process like the motion of the earth but any arbitrary motion of the same kind. In case of the planetary motion a law with this degree of generality is given by Newton's equation of motion of a body in the gravitational field of the sun. The trajectory of each planet can then be determined by Newton's law together with convenient initial conditions. The more special laws discovered by Kepler can then be derived – in a convenient approximation – from the general Newtonian law. Hence a law is an abstraction that describes the observed regularities apart from concrete processes and specific objects. (cf. 11.2)

Another important requirement, which must be fulfilled by a law of nature, is universality. A law should hold not only at a certain instant of time but always and not only at a particular place but everywhere. In should hold in all equivalent situations, where the concept of "equivalence" must explicitly be explained. For example, in Newtonian mechanics all systems of inertia are dynamically equivalent. Requirements of this kind are usually expressed by invariance postulates, which must be fulfilled by laws of nature. In addition, a law of this kind must be compatible with other laws such that various at first

glance independent laws can be comprehended in a "theory", a formal system that allows to derive new consequences describing new experimental facts.

If a particular law or a network of laws, which we call "theory", fulfils the mentioned requirements of universal validity, then the question arises *why* the laws considered hold. Does the claim to universal validity indicate a hidden, not yet recognised *necessity* of the observed regularities? We will follow this question here and ask at first for possible nonempirical reasons and justifications for a law of nature.

12.1.2 Why are Laws of Nature Valid?

We mention here three possible justifications for the necessity of a law of nature:

(a) A law of nature could hold necessarily since it is merely a disguised law of logic or of mathematics. In Chap. 4 it became obvious that – irrespective of their mathematical justification – the laws of arithmetic, geometry of the three-dimensional space, and probability could also be considered as laws of nature. They are applicable to the physical reality and they hold rigorously in all conceivable situations. Hence they fulfil the necessary requirements of laws of nature but their validity is not based on experience but on mathematical reasoning.

Another example is quantum logic, which will be discussed in Chap. 13. The universal empirical validity of quantum mechanics will not be questioned here. However, we query the empirical justification of quantum mechanics and ask for an additional argument that demonstrates the necessity of quantum mechanics. It turns out that the formal framework of quantum mechanics, the Hilbert space with complex numbers, can be traced back to the orthomodular lattice of its subspaces and that this lattice is the Lindenbaum–Tarski algebra of the calculus of quantum logic. Since all laws of quantum logic are also valid in ordinary classical logic and since quantum logic follows from the most general conditions of a formal language of physics we find that the validity of quantum mechanics is not only based on experience but that this theory can be justified by purely logical and mathematical reasoning.

(b) A law of nature holds necessarily for a certain real entity – an object, a property, or a process – if the validity of the law in question is a necessary precondition of that entity. In Sects. 10.2 and 10.3 we demonstrated that for the constitution of objects in classical mechanics and in quantum mechanics invariance principles and laws of physics must be presupposed. In particular it turned out that observables and symmetry transformations must fulfil certain covariance postulates. Hence it is obvious that the laws and symmetries that were used for the constitution of objects hold necessarily for these objects. This is the case in particular for the invariance properties in question. Although this result is almost trivial it demonstrates a further kind of necessary laws of nature.

The way of reasoning, which is applied here, is not new but can be traced back to the transcendental arguments in the philosophy of *Kant.* In the *Critique of Pure Reason* Kant emphasised that "objects of experience" cannot directly be recognised but must be constructed, composed, or constituted on the basis of sensations and observations by means of conceptual prescriptions, the categories of substance and causality. This means that our impressions and sensations must be ordered and interpreted by these categories in order to obtain objects, i.e. entities of the external reality. For these constituted objects the laws of causality and conservation of substance hold a priori, i.e. independent of and prior to any possible experience. However, these arguments don't show the necessity of a particular law of nature, but only the necessary validity of the very general rules of causality and conservation of substance. Kant's attempts to extend his results in a similar transcendental way to some laws of classical Newtonian mechanics[1] are not convincing and will thus not be considered in the present investigations.

(c) There is a third, completely different way for demonstrating that a law of nature holds necessarily. For compound systems which are composed of a very large number of subsystems it can happen, that the compound system obeys laws which follow exclusively from the large number of the single systems, i.e. laws which are based merely on the statistical law of large numbers. This means that the laws that govern the individual subsystems have no influence on the laws that determine the behaviour of the ensemble. This possibility was already indicated by Schrödinger in his inaugural lecture of 1922 (cf. Sect. 7.2.3.4.2)[2] and further elaborated by Wheeler who called a statistical law that holds without any recourse to individual laws a "law without law".[3]

Hence it could happen that some laws of statistical mechanics hold irrespective of the laws that govern the individual particles, i.e. the laws of classical Newtonian mechanics. However, a test of this conjecture is a difficult task, since one had to show that also alternative non-Newtonian laws for the individual particles would lead to the well known laws of statistical mechanics. The problem behind this way of reasoning is that alternative theories for classical mechanics could be inconsistent in many ways. Presently we don't know why the laws of Newtonian mechanics hold and whether modifications of these laws are free from internal inconsistencies.

Within the domain of statistical mechanics and thermodynamics one can, however test a somewhat weaker conjecture, which seems to be sufficient for our problem. We could presuppose, that individual systems are governed

[1] The most interesting exemples of this kind can be found in *the Metaphysische Anfangsgründe der Naturwissenschaft;* further attempts are in Kant's *Opus Postumum.* (cf. Vittorio Mathieu (1989, KOP))

[2] cf. Schrödinger (1961, WNG) p.11

[3] Wheeler (1983, RLL)

merely by very general and less specific laws like invariance principles and derive from these weak assumptions a law of statistical thermodynamics. This will be done in Sect. 12.3.1 where Boltzmann's energy distribution law is derived from very weak assumptions.

It would be even better, if we could find a field of physics where individual objects are not determined by any law. If under these extreme conditions statistical laws for a large ensemble can be derived then these laws hold necessarily. In the domain of quantum mechanics we find situations of this kind. In Sect. 12.3.2 we will show that the probabilistic laws of quantum mechanics can be derived although the single quantum systems are not determined by any law of nature. Hence, these probabilistic laws of quantum mechanics are necessary.

The question posed at the beginning of the present section, whether there are laws of nature that hold necessarily, can thus be answered in the affirmative. However, this does *not* answer the question whether in addition to these laws there are also contingent laws, i.e. laws that cannot be justified by a priori reasoning. We must leave this question open in particular since it is not clear what it means that a law – not a mere regularity – is contingent.

12.2 Are Statistical Laws Based on Individual Laws?

In classical statistical physics there are strong indications that statistical laws which hold for macroscopic systems, i.e. for large ensembles of microscopic objects are based on dynamic laws and properties of the single objects considered. (cf. Chap. 6) Usually, we assume that the single objects are governed by the laws of Newtonian mechanics, electrodynamics, and gravity and that processes like collisions, motion in an electromagnetic or gravitational field etc. are adequately described by them. Many statistical laws follow from these premises. We mention here the relation between the temperature of an ideal gas of molecules and the kinetic energy of the single molecules, the barometer formula which connects the density of a gas as function of the altitude with the motion of molecules in the gravitational field of the earth, and many other thermodynamic relations which can be traced back to the dynamics of electrons, atoms, and molecules.[4]

In this way a dualistic picture of statistical physics and classical dynamics of single objects seems to be well established. The behaviour of macroscopic quantities like temperature, density, entropy etc. is determined by the dynamic behaviour of the underlying microscopic objects like electrons, atoms, and molecules. However, these results do not show that for every law which connects statistical quantities corresponding laws of individual objects can be found which induce the statistical laws mentioned. Of course, outside of

[4] cf. for example the textbook on *Statistical Physics* by Landau and Lifschitz (1966, ThP)

physics there are many statistical regularities which are not justified on the basis of individual laws but simply based on empirical evidence. We mention here the statistical results used in life insurance, or the statistics about relative frequencies of car accidents etc. These problems are outside the scope of our consideration. Instead, we are interested in the question whether there are well-confirmed statistical *laws of nature* that are not based on dynamical laws for the single systems. The answer to this question requires a more detailed investigation. We will proceed here in two steps, first discussing problems of classical statistics and second problems of quantum statistics.

12.3 Are there Statistical Laws without Individual Laws?

Arguments Contra

(1) If there were statistical laws of an ensemble of molecules or atoms, say, without individual laws, then there would be no laws governing the single molecules and atoms. In this case, the single molecules and atoms would not even be guided by the conservation laws (like that of energy, charge, etc.). But the conservation laws are assumed to hold universally and they are the most well established laws of physics. For the conservation of energy this argument is confirmed by the example in Sect. 12.3.1 on classical statistics.

Therefore, it is impossible that there are statistical laws without individual laws.

(2) If there were statistical laws without individual laws then there would be no laws at all governing the single elementary particles. Hence, the single elementary particles could not be guided by dynamical laws. But as a matter of fact – and without any exception – the single particles obey the law of gravitation, since gravity is a universal force.

Therefore, it is impossible that there are statistical laws without individual laws.

Arguments Pro

(1) If for every individual system, say an individual neutron or photon (in a quantum mechanical measurement) the value of the measured observable is objectively undetermined before the measurement, whereas a sufficiently large number of photons (or neutrons) satisfy a statistical law, then there are statistical laws without individual laws. Indeed, recent split-beam experiments show that there is indeterminacy in a strict objective sense for every individual system and yet an emergence of a statistical law (if the ensemble is large enough) which provides an objective and definite property of the whole system. In quantum mechanics, this result holds also for

conserved quantities like energy or momentum and for gravity induced phenomena as the path of a particle.[5] In this sense Wheeler spoke of "law without law"

"All physics, in my view, will be seen someday to follow the pattern ... of regularity based on chaos, of '*law without law*'."[6]

Therefore, there are statistical laws without individual laws.

12.3.1 Classical Statistics

The first example for a "law without law" which we will discuss here is taken from statistical mechanics. The behaviour of a large ensemble of molecules is governed by Boltzmann's law:[7]

The probability p for a molecule to be in a state of energy ε is given by the formula

$$p(\varepsilon) = p_0(N)e^{-\varepsilon/\gamma}$$

where $p_0(N)$ is a normalising factor depending on the number N of molecules, ε is the energy of the molecular state in question and γ is the mean energy of a molecule which is proportional to the temperature. Hence, we are confronted with the important question "how can stupid molecules ever be conceived to obey a law so simple and so general?" (Wheeler (1983, RLL)). The predicate "stupid" means in this context that the molecules are not governed by any law.

In order to answer this question let us consider N oscillators and K quanta of energy. By $\mu(K, N)$ we denote the number of ways of sharing the energy (K quanta) between N oscillators. E.g. for $N = 2$ oscillators we have $K + 1$ ways to distribute K quanta according to the scheme:

$$[0 \,|\, K], \qquad [1 \,|\, K-1], \qquad [2 \,|\, K-2], \ldots, [K \,|\, 0]\ .$$

In the general case of N oscillators and K quanta of energy the number of ways to distribute the energy is given by

$$\mu(K, N) = (K + N - 1)!/K!(N - 1)!\ .$$

This formula can be further elucidated, if we consider the number $\nu_K(x; N)$ of possibilities to have x quanta in one oscillator and $K - x$ quanta in the remaining $N - 1$ oscillators. This number is given by

$$\nu_K(x; N) = \mu(K - x, N - 1) = (K - x + N - 2)!/(K - x)!(N - 2)!\ .$$

For the special case $K = 4, N = 4$ the graph of this formula (Fig. 12.1) shows a rapidly decreasing function.

[5] This can be illustrated by the split-beam experiment of Sect. 12.2.2 with neutrons in a gravitational field which was realised by Rauch et al. cf. Rauch (1988,NIT)

[6] Wheeler (1983, RLL), p. 398

[7] Boltzmann (1868, SKA)

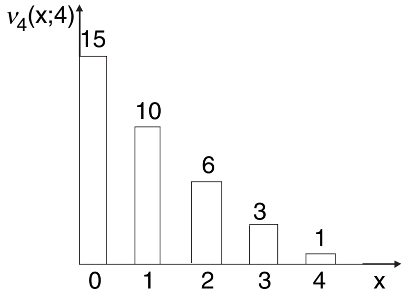


Fig. 12.1. The number $v_4(x;4)$ of possibilities to have x quanta in one oscillator

With increasing numbers of N and K of oscillators and quanta, respectively, the formula shows more and more exponential behaviour. For $K = N = 6$ we obtain the graph in Fig. 12.2 which confirms this conjecture.

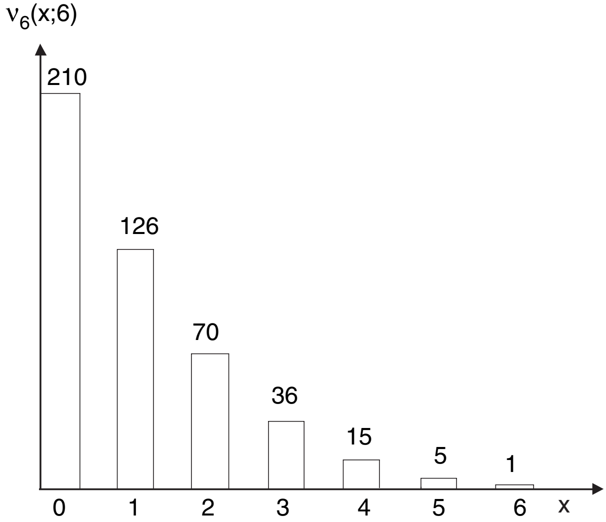


Fig. 12.2. The number $v_6(x;\,6)$ of possibilities to have x quanta in one oscillator

In the limit of very large numbers N and K one obtains Boltzmann's formula. This can be shown in the following way: The number of possibilities to have x quanta of energy ε_0, i.e. the energy $\varepsilon = x \cdot \varepsilon_0$ in one oscillator is given by

$$\mu(K - x, N - 1) = \{1/(N - 2)!\} \prod_{i=1}^{N-2} (K - x + N - 1 - i) \,.$$

Here we write $E = K \cdot \varepsilon_0$ for the total energy and $\gamma := (E - \varepsilon)/(N-1)$ for the mean energy of one of the remaining $(N-1)$ oscillators. Hence $K - x = (E - \varepsilon)/\varepsilon_0$, and thus we find

$$\begin{aligned} K - x + N - 1 - i &= (E - \varepsilon)/\varepsilon_0 + N - 1 - i \\ &= \left(\frac{E + (N-1-i)\varepsilon_0}{\varepsilon_0}\right)\left(1 - \frac{\varepsilon}{E + (N-1-i)\varepsilon_0}\right) \end{aligned}$$

and

$$\begin{aligned} &\mu(K - x, N - 1) \\ &= \{1/(N-2)!\} \prod_{i=1}^{N-2} \frac{E + (N-1-i)\varepsilon_0}{\varepsilon_0} \quad \prod_{i=1}^{N-2} 1 - \frac{\varepsilon}{E + (N-1-i)\varepsilon_0} \end{aligned}$$

We will now make the following approximations

$$E \gg \varepsilon, K - x \gg N - 1 \gg 1 \,.$$

It therefore follows $\gamma \gg \varepsilon_0$ and thus

$$E + (N-1-i)\varepsilon_0 \approx (N-1)\gamma + (N-1-i)\varepsilon_0 \approx (N-1)\gamma$$

for all i. The equation for $\mu(k - x, N - 1)$ now becomes

$$\mu(k - x, N - 1) = \frac{1}{(N-2)!}((N-1)\gamma/\varepsilon_0)^{N-2}(1 - \varepsilon/(N-1)\gamma)^{N-2}$$

Using the relation

$$e^{-\lambda} = \lim_{s\to\infty}(1 - \lambda/s)^s$$

for $N \gg 1$ we find

$$\mu(K - x, N - 1) = C(N)\, e^{-\varepsilon/\gamma}$$

with $C(N) = [1/(N-2)!]((N-1)\gamma/\varepsilon_0)^{N-2}$. Finally, normalisation leads to the Boltzmann formula mentioned above

$$p(\varepsilon) = p_0(N)e^{-\varepsilon/\gamma}\,.$$

Where does this formula come from? For the derivation there is no need to presuppose dynamical laws which are fulfilled by oscillators or quanta of energy. There is only one restriction: The total amount of energy $E = \varepsilon_0 K$ contained in K quanta of energy must be constant. Hence, Boltzmann's formula is not exactly a "law without law" but a law that follows from a weak law-like assumption, the conservation of energy. This energy conservation law, which corresponds to an invariance property, is a genuine law of nature. Hence, Boltzmann's law – though very surprising at first glance – is finally based on an empirical law, which can, in principle, be falsified. This means that the "stupid molecules" in Wheelers above-mentioned question are not completely stupid. They obey the law of conservation of energy.

12.3.2 Quantum Statistics

Our second example for a "law without law" is taken from quantum mechanics. Here, we are confronted with a very particular situation. The laws that are primarily provided by the theory are statistical laws and these laws are confirmed experimentally to a very high degree of accuracy. However, for a single system we do not know a dynamical law – like Newton's equation of motion – that determines the properties of the system considered. Moreover, in quantum mechanics it is not possible to assume hypothetically that the behaviour of an individual system is determined by a dynamical law that we do not know. Hence, in quantum physics a situation is realised such that we have well-established statistical laws but objective indeterminacy for individual systems. The statistical laws are laws without a law-like background and we will ask where these laws come from.

Objective Indeterminacy versus Statistical Determination

For illustrating this situation we consider the split beam experiment in Fig. 12.3, which was realised both with photons and neutrons. In this experimental set-up the state φ of the incoming photon is split by a half-transparent mirror, beam splitter BS_1, into two orthogonal components described by orthonormal states φ^B and $\varphi^{\neg B}$. The two parts of the split beam are reflected at two (fully reflecting) mirrors M_1 and M_2 and recombined with a phase shift δ at a second half-transparent mirror, beam splitter BS_2. In the experiment there are two mutually exclusive measuring arrangements: If the detectors D_1

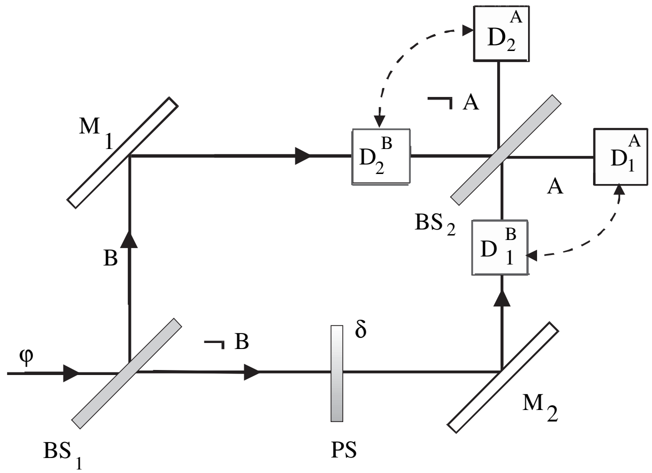


Fig. 12.3. Photon split-beam experiment with beam splitters BS_1 and BS_2, both half-reflecting mirrors, two fully reflecting mirrors M_1 and M_2, a phase shifter PS providing a phase shift δ, and two detectors D_1 and D_2 in mutually exclusive positions $({D_1}^A, {D_2}^A)$ and $({D_1}^B, {D_2}^B)$

and D_2 are in the positions (D_1^B, D_2^B) one observes which way (B or $\neg B$) the photon came. If the detectors are in the position (D_1^A, D_2^A) one observes the interference pattern, i.e. the intensities which depend on the phase δ.

In the latter case of position (D_1^A, D_2^A) let us first consider the special case $\delta = 0$ without phase shift. In this particular situation the outcomes of the split-beam experiment is completely determined. destructive interference destroys all radiation going to the second detector D_2^A, whereas the first detector D_1^A registers, in case of ideal sensitivity, every incoming photon. However, even in this situation of fully determined outcomes nothing can be said about the path of a single photon, neither which way the photon came nor that it travelled both routes.

In the general case of a non-vanishing phase shift $\delta \neq 0$ for an individual photon it is no longer determined whether it will be registered at detector D_1 or D_2. However, in spite of the indeterminacy in each individual case, for a large number of identically prepared photons the preparation φ and the observable A induce a probability for A (to register the photon in D_1^A) and for $\neg A$ (to register the photon in D_2^A) which reads

$$p(\varphi, A) = \cos^2 \delta/2 \qquad \text{and} \qquad p(\varphi, \neg A) = \sin^2 \delta/2,$$

respectively. This means that the relative frequency of photons arriving at D_1^A is approximately given by $\cos^2 \delta/2$ and the relative frequency of systems arriving at D_2^A by $\sin^2 \delta/2$. Under these conditions one is tempted to assume that for each individual photon the state or the value after measurement can be attributed to the object even before the measurement, although these data are not yet known to the observer in this situation.

It is one of the most important results of quantum mechanics that this assumption is not correct. Neither an eigenstate of A nor the corresponding eigenvalue can be attributed to the photon system in the preparation state φ before the measurement of A. Indeed, if the observable A could be weakly objectified to the system S in state φ, i.e. if an A-value could be attributed to $S(\varphi)$, then the system S were in the mixed state.[8]

$$W_S(\varphi, A) = p(\varphi, A)P[\varphi^A] + p(\varphi, \neg A)P[\varphi^{\neg A}]$$

and not in the pure state $P[\varphi]$. Here we have denoted the two eigenstates of A by φ^A and $\varphi^{\neg A}$, respectively. However, it can easily be shown that the states $P[\varphi]$ and W_S (φ, A) are not generally equivalent.

In order to demonstrate this important result one must compare the probabilities of a convenient test-observable C for the two situations mentioned, i.e. for the pure state $P[\varphi]$ and for the mixed state W_S (φ, A). Experimentally, this comparison can be performed by extending the beam-split experiment such that the two beams of A and $\neg A$, respectively, are recombined with a phase shift ε by a third half-transparent mirror BS_I (Fig. 12.4). The photons

[8] Busch, P. et al. (1992,WOB)

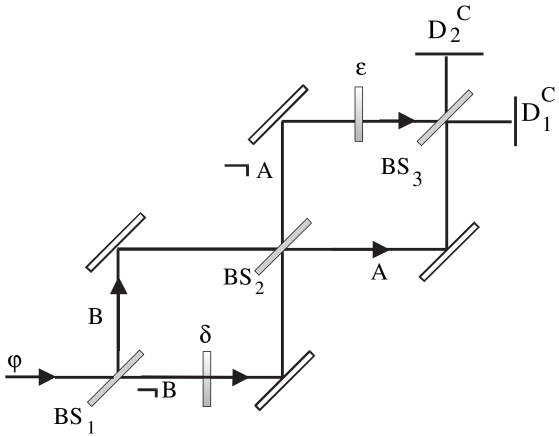


Fig. 12.4. Extended split beam experiment with three half-reflecting mirrors BS_1, BS_2, and BS_3, two phase shifters providing phase shifts δ and ε, and two Detectors D_1 and D_2 in positions $(D_1^C,\, D_2^C)$

will then be registered by the detectors in positions $(D_1^C,\, D_2^C)$. The probability to register a photon in the detector D_2^C is then given for the pure state P[φ] by

$$p(\varphi, D_2) = \frac{1}{2}(1 + \sin\delta \sin\varepsilon)$$

and for the mixed state by

$$p(W_S(\varphi, A), D_2) = \frac{1}{2}$$

Clearly, the probability $p(\varphi,\, D_2)$ shows interference pattern in accordance with experiments, which could easily be performed, whereas $p(W_S(\varphi, A), D_2)$ does not depend on ε at all and thus does not show any interference pattern. Hence, we find, that weak objectification of A in the state φ leads to predictions about the statistics of measurement outcomes which are incompatible both with quantum theory and with realisable experiments. For this reason weak objectification of A in the state φ is not possible. This means that the result of an A-measurement performed at an individual photon is not only subjectively unknown for the observer but objectively undecided.

Summarising the discussion of this section we arrive at the following two at the first glance incompatible results: On the one hand, for each individual photon the value of the observable A is objectively undetermined prior to the registration in one of the detectors. On the other hand, a sufficiently large number of photons fulfils the statistical law which states that the relative frequency of photons registered in detector D_1^A is approximately given by $\cos^2 \delta/2$. Hence, we find that a large number of lawless quantum systems fulfils a strict probability law. Under these conditions one may wonder where

this statistical law comes from. Again, we will pose here Wheeler's question whether we are confronted here with a "law without law", i.e. a law-like behaviour of a large number of elements which emerges without any causal reason. Hence, we will investigate the question whether the quantum probability law emerges by measuring a large ensemble of lawless single quantum systems.

The Quantum Mechanical Measurement Process

The split beam experiment of the proceeding section can be described in a two dimensional Hilbert space $H_2 = C^2$ which is conveniently illustrated by the Poincare' sphere P (Fig. 12.5). If we denote the orthogonal unit vectors of the sphere by $\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{n}_3$ and make use of the Pauli matrices $\sigma_1, \sigma_2, \sigma_3$ then the projection operators in $H_2 = \mathrm{C}^2$ can be written as

$$P(\mathbf{n}) = \frac{1}{2}\left(\mathbf{1} + n_i \sigma_i\right) \qquad n_1^2 + n_2^2 + n_3^2 = 1$$

where $\mathbf{n}= (n_1, n_2, n_3)$ is a unit vector. Hence the projection operators correspond to points on the surface of the Poincare' sphere. The properties B and A of the preceding section can then be expressed by the projection operators

$$P(B) := P(\mathbf{n}_3) = \frac{1}{2}(\mathbf{1} + \sigma_3), P(\neg B) := P(-\mathbf{n}_3)$$

with eigenstates φ^B and $\varphi^{\neg B}$ and

$$P(A) := P(\mathbf{n}_1) = \frac{1}{2}(\mathbf{1} + \sigma_1),\ P(\neg A) := P(-\mathbf{n}_1)$$

with eigenstates φ^A and $\varphi^{\neg A}$. The preparation of the system φ of the system (photon, neutron, etc.) is then given by $\varphi = 1/\sqrt{2}\ (\ \varphi^B + e^{i\delta}\varphi^{\neg B})$ or by the projection operator

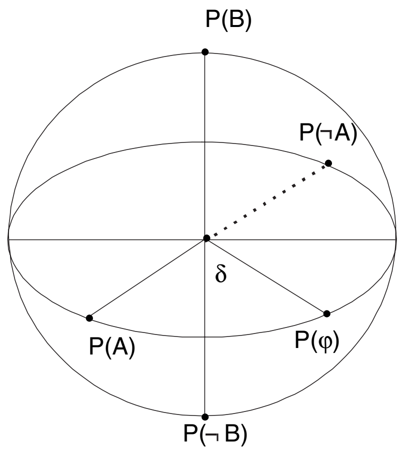


Fig. 12.5. Poincaré sphere of the photon split-beam experiment shown in Fig. 12.3

$$P[\varphi] = P(\varphi) = \frac{1}{2}(\mathbf{1} + \cos\delta\sigma_1 + \sin\delta\sigma_2) \tag{1}$$

with $\varphi = (\cos\delta,\ \sin\delta, 0)$ and corresponds to a point an the equator of P. The angle between vectors φ and $\mathbf{n}_1$ is then given by the phase shift δ.

For the probabilities of B and A we obtain

$$\begin{aligned} p(\varphi, B) &= tr\{P[\varphi]P(B)\} = \frac{1}{2}(1 + \varphi \cdot \mathbf{n}_3) = {}^1\!/_2 \\ p(\varphi, A) &= tr\{P[\varphi]P(A)\} = \frac{1}{2}(1 + \varphi \cdot \mathbf{n}_1) = \cos^2\delta/2 \end{aligned} \tag{2}$$

In order to describe the measuring process of the observable $A : P(A)$ we consider the object system S with Hilbert space H_S and the measuring apparatus M with Hilbert space H_M. Let $\varphi \in H_S$ and $\Phi \in H_M$ be the preparations of S and M, respectively, i.e. the states prior to the measurement process. Here we consider a unitary and repeatable premeasurement of A, which can be described by a unitary operator U_A acting on the tensor product $\varphi \otimes \Phi$ of the compound system $S + M$. The operator U_A is further determined by the calibration postulate. If the object system S is in one of the two eigenstates φ^A or $\varphi^{\neg A}$ of A, then the unitary and repeatable premeasurement must reproduce this state. This means that for the special preparation $\varphi(\delta = 0) = \varphi^A$ and $\varphi(\delta = \pi) = \varphi^{\neg A}$ we have

$$\begin{aligned} \varphi^A \otimes \Phi &\to U_A(\varphi^A \otimes \Phi) = \varphi^A \otimes \Phi_A \\ \varphi^{\neg^A} \otimes \Phi &\to U_A(\varphi^{\neg A} \otimes \Phi) = \varphi^{\neg A} \otimes \Phi_{\neg A} \end{aligned} \tag{3}$$

where Φ_A and $\Phi_{\neg A}$ are eigenstates of a pointer observable $Z = Z_A\ P[\Phi_A] + Z_{\neg A} P[\Phi_{\neg A}]$ whose eigenvalues Z_A and $Z_{\neg A}$ indicate the measuring results A and ¬A, respectively. From (2) it follows by the unitarity of U_A for an arbitrary preparation $\varphi = (\varphi^A, \varphi)\varphi^A + (\varphi^{\neg A}, \varphi)\varphi^{\neg A}$

$$\varphi \otimes \Phi \to U_A(\varphi \otimes \Phi) = (\varphi^A, \varphi)\varphi^A \otimes \Phi_A + (\varphi^{\neg A}, \varphi)\varphi^{\neg A} \otimes \Phi_{\neg A}\ .$$

For the particular preparation φ given by (1) we find for the state of $S + M$ after the measurement

$$U_A(\varphi \otimes \Phi) = \frac{1}{2}(1 + e^{i\delta})\varphi^A \otimes \Phi_A + \frac{1}{2}(1 - e^{i\delta})\varphi^{\neg A} \otimes \Phi_{\neg A}\ .$$

The state of the object system after the premeasurement is then given by the reduced mixed state

$$W_S(\varphi, A) = \cos^2\delta/2\ P[\varphi^A] + \sin^2\delta/2\ P[\varphi^{\neg A}]\ .$$

The interpretation of this mixed state of the object system after the premeasurement is usually given by the probability reproducibility condition: The probability distribution $p(\varphi, A_i), A_i \in \{A, \neg A\}$ which is induced by the

preparation φ and the measured observable $P(A)$, is reproduced in the statistics of the post-measurement values $(Z_A, Z_{\neg A})$ and states $(\varphi_A, \varphi_{\neg A})$ of the pointer. In case of repeatable measurements this means that $p(\varphi, A_i)$ is also reproduced in the statistics of the states $(\varphi^A, \varphi^{\neg A})$. On the basis of these arguments the main question of this paper can now be formulated as follows: Given an ensemble of (before the measurement) identically prepared systems S that are in the reduced state

$$W_S(\varphi, A) = p(\varphi, A)P[\varphi^A] + p(\varphi, \neg A)P[\varphi^{\neg A}]$$

after the premeasurement of A. Is it possible to justify that the (formal) probability $p(\varphi, A_i)$ is reproduced in the statistics of the measurement results A_i? A justification of this kind could presumably explain in which way a large number of objectively undetermined objects fulfil a strict statistical law.

The Probability Reproducibility Condition

Consider a large number of identically prepared systems S_i in states $\varphi^{(i)}$ which are not eigenstates of the observable A. Let us further assume that the unitary operator U_A, which is used for a measurement of the observable A, fulfils the calibration postulate for repeatable measurements. Then we know that a measurement of the observable A in case of the particular preparation φ^A leads with certainty to the states Φ_A and φ^A showing the result A. On the basis of this probability free interpretation of quantum mechanics we want to show that for arbitrary preparations $\varphi \neq \varphi^A, \varphi \neq \varphi^{\neg A}$ the formal probability $p(\varphi, A_i)$, which is induced by φ and A, is reproduced in the statistics of the measuring outcomes A_i. The probability reproducibility condition would then be a theorem of the probability free theorem and no longer an additional postulate.

Let us consider N independent systems S_i with identical preparation $\varphi^{(i)}$ as a compound system $S^{(N)}$ in the tensor product state

$$(\varphi)^N = \varphi^{(1)} \otimes \varphi^{(2)} \cdots \otimes \varphi^{(N)}, \quad (\varphi)^N \in H_S^{(N)}$$

where $H_S^{(N)}$ is the tensor product of N Hilbert spaces $H(S_i)$. A premeasurement of A transforms the initial state $\varphi^{(i)}$ of each system S_i into the mixed state

$$W'_S = \sum p(\varphi, A_i)P[\varphi(A_i)]$$

with eigenstates $\varphi(A_i)$ of A corresponding to values A_i. If A is measured on each system S_i, then the measuring result is given by a sequence $\{A_{l(1)}, A_{l(2)}, \cdots A_{l(N)}\}$ of system values $A_{l(i)}$ and states $\varphi(A_{l(i)})$, respectively, with an index sequence l: $= \{\mathrm{l}(1), \mathrm{l}(2), \cdots l(N)\}$ such that $A_{l(i)} \in \{A_k\}$.

In the N-fold tensor product Hilbert space ${H_S}^{(N)}$ of the compound system $\mathrm{S}^{(N)}$ the special states $(\varphi)_l^N = \varphi_{l(1)}^{(1)} \otimes \cdots \otimes \varphi_{l(N)}^{(N)}$ with $\varphi_{l(i)}^{(i)} := \varphi\ (A_{l(i)}) \in H(S_i)$ form an orthonormal basis. The relative frequency $f^N(k, l)$ of values A_k in the

state $(\varphi)_l^N$ is then given by $f^N(k,l) = 1/N \sum \delta_{l(i),k}$. We can then define in $H_S^{(N)}$ an operator "relative frequency of systems with values A_k" by

$$f_k^N := \sum f^N(k,l)P[(\varphi)_l^N]$$

where the sum runs over all sequences l. The eigenvalue equation of this operator

$$f_k^N(\varphi)_l^N = f^N(k,l)(\varphi)_l^N$$

then shows that the relative frequency of the value A_k is an objective property of $\mathrm{S}^{(N)}$ in the state $(\varphi)_l^N$ and given by $f^N(k,l)$. The eigenvalue equation can also be written in the equivalent form

$$tr\{P[(\varphi)_l^N](f_k^N - f^N(k,l))^2\} = 0 \ .$$

After a premeasurement of A a system S_i is in a mixed state W_S'. If N premeasurements of A are performed, then the state of the compound system $S^{(N)}$ is given by the N-fold tensor product state

$$(W_S')^N = W_1' \otimes W_2' \otimes \cdots \otimes W_N'$$

of these mixed states W_i'. One easily verifies that the expectation value of f_k^N in this product state is given by

$$tr\{f_k^N(W_S')^N\} = p(\varphi, A_k) \ .$$

However, in general the state $(W_S')^N$ is not eigenstate of the relative frequency operator f_k^N. This means that

$$T_k^N := tr\{(W_S')^N(f_k^N - p(\varphi, A_k))^2\} \neq 0$$

and that the relative frequency of values A_k is not an objective property of the system $S^{(N)}$ in the state $(W_S')^N$.

In contrast to this somewhat unsatisfactory result one finds that for large values of N the post-measurement product state $(W_S')^N$ of the compound system $S^{(N)}$ becomes an eigenstate of the operator f_k^N and the value of the relative frequency of A_k approaches the probability $p(\varphi, A_k)$. Indeed, one finds after some tedious calculations[9,10]

$$T_k^N = 1/Np(\varphi, A_k)(1 - p(\varphi, A_k))$$

and thus one finally obtains the desired result

$$\lim_{N\to\infty} tr\{(W_S')^N(f_k^N - p(\varphi, A_k)^2\} = 0$$

[9] DeWitt (1971, MUI)

[10] Mittelstaedt (1990, OMI), (1993, MPI), (1998, IQM)

This means that in the limit of an infinite number N of systems the state $(W'_S)^N$ is an eigenstate of the operator f_k^N of the relative frequency of values A_k and that the compound system $S^{(N)}$ possesses the relative frequency $p(\varphi, A_k)$ of A_k as an objective property.

In the present example of a photon split-beam experiment there are two possible outcomes $A_1 = A$ and $A_2 = \neg\, A$ with probabilities $p(\varphi, A_1) = \cos^2 \delta/2$ and $p(\varphi, A_2) = \sin^2 \delta/2$. The decomposition of the preparation φ into eigenstates $\varphi(A_i)$ reads

$$\varphi = \frac{1}{2}(1 + e^{i\delta})\varphi(A) + {}^1/_2(1 - e^{i\delta})\varphi(\neg A)$$

and the post-premeasurement mixed state is given by

$$W'_S = \cos^2 \delta/2 \; P[\varphi\,(A)] + \sin^2 \delta/2 \; P[\varphi\,(\neg A)]$$

For finite N the expectation value of the relative frequency operator in the state $(W'_S)^N$ is always given by

$$tr\{(W'_S)^N f_k^N\} = p(\varphi, A_k) =: p_k(\delta)$$

but $(W'_S)^N$ is not an eigenstate of f_k^N. However, for the expression $T_k^N(\delta)$ we obtain

$$T_k^N(\delta) = tr\{(W'_S)^N (f_k^N - p_k(\delta))^2\} = (1/4N) \sin^2 \delta$$

which shows that for increasing N the mixed state $(W'_S)^N$ becomes an eigenstate of f_k^N.

In order to ensure this way of reasoning against mathematical objections one has to guarantee that the overwhelming majority of index sequences $l = \{l(i)\}$ are random sequences and that the contribution of the non random sequences can be neglected. As a first orientation let us define the function $\delta(l) = \sum_k (f^N(k, l) - p(\varphi, A_k))^2$ in order to measure the degree to which a given sequence l deviates from a random sequence with weights $p(\varphi, A_k)$. A sequence l will be called first random if $\delta(l) < \varepsilon$ for an arbitrary positive ε. In the limit N$\rightarrow \infty$ the contribution of the non first random sequences disappear. This can be shown in the following way:[11]

A unitary premeasurement (of the observable A) of one system S_i in the state $\varphi^{(i)}$ with an initial pointer state $\Phi^{(i)}$ leads to the $S + M$ compound state

$$\Psi' = U_A(\varphi^{(i)} \otimes \Phi^{(i)}) = \sum_k c_k \varphi_k^{(i)} \otimes \Phi_k^{(i)}$$

with pointer eigenstates $\Phi_k^{(i)}$ and coefficients $c_k = (\varphi_k^{(i)}, \varphi^{(i)})$. If a measurement process is performed with N equally prepared systems S_i we have to consider the final state

[11] Mittelstaedt (1990, OMI), Busch et al. (1996, QTM), (pp.48, 49)

$$\Psi^{(N)'} = \otimes_{i=1}^{N} \left(\sum_k c_k \varphi_k^{(\mathrm{i})} \otimes \Phi_k^{(i)} \right) = \sum_l c_{\{l\}} (\varphi)_l^N \otimes \Phi_l^N$$

with index sequences $l = \{l(i)\}$, coefficients $c_{\{l\}} = c_{l(1)} \cdot c_{l(2)} \cdot \ldots c_{l(N)}$ and pointer eigenstates tensor products $\Phi_l^N = \Phi_{l(1)}^{(1)} \otimes \Phi_{l(2)}^{(2)} \ldots \otimes \Phi_{l(N)}^N$. If we remove from the superposition $\Psi^{(N)'}$ all sequences l that are not first random, i.e. all l with $\delta(l) \geq \varepsilon$, then we obtain

$$\Psi_\varepsilon^{(N)'} = \sum_{l:\delta<\varepsilon} c_{\{l\}} (\varphi)_l^N \otimes \Phi_\varepsilon^N$$

and the difference between $\Psi^{(N)'}$ and $\Psi_\varepsilon^{(N)'}$ is given by

$$\chi_\varepsilon^N := \Psi^{(N)'} - \Psi_\varepsilon^{(N)'} = \sum_{l:\delta\geq\varepsilon} c_{\{l\}} (\varphi)_l^N \otimes \Phi_l^N .$$

Using some results for the relative frequency $f^N(k,l)$ one finds that

$$(\chi_\varepsilon^N, \chi_\varepsilon^N) \leq \frac{1}{N\varepsilon} \sum_k p(\varphi, A_k)[1 - p(\varphi, A_k)] \leq \frac{1}{N\varepsilon}$$

which means that in the limit $N \to \infty$ the contribution of the non-first-random sequences becomes arbitrary small. This first confirmation of our statistical results will be made more rigorous in the following section.

Mathematical Considerations

In classical probability theory it is well known that probabilities are not relative frequencies and that relative frequencies are not probabilities.[12] For a given experimental process, e.g. dice tossing, the probability for a certain result can be determined by mathematical means only. The interpretation of probabilities by relative frequencies suggests that in a sequence of ideal dice tossing events the relative frequency of a certain result, five spots, say, approaches the calculated probability 1/6 provided the number N of events is sufficiently large. This is, however, not entirely correct. There are counter examples, which cannot be excluded. For example, the sequence that consists only of results with three spots is not forbidden by any law of nature. However, sequences of this kind, the non-random sequences, are very rare compared to the normal random sequences. More precisely, within the set of all sequences of events the subset of non-random sequences is of measure zero. This is the content of the "law of large numbers": For large numbers N the relative frequency $f^N(k,l)$ for some index value k approaches the probability $p(k)$ for almost all sequences l, i.e. with a probability which is equal to one.[13] This

[12] This was emphasised in particular by R. v. Mises (1931, WAS)

[13] For details cf. Richter (1956, WTh), pp. 52–57 and Bauer (1978, WGH), pp. 165–71.

means that probability statements cannot be replaced in general by probability free statements, even if the number of samples is infinite. Hence, it must be clarified for the above-mentioned result that in quantum mechanics probabilities can in fact be completely eliminated, provided the number of systems or of measurement outcomes is infinite.

According to a recent investigation by Gutman[14] for the problem of the present paper this can actually be shown. Let $S^{(\infty)}$ be a compound system which is composed of infinitely many copies of split beam photons S_i. The system a $S^{(\infty)}$ can be described in the infinite tensor product space $H_S^{(\infty)} = H(S_1) \otimes H(S_2) \cdots$ which is a nonseparable Hilbert space.[15] The eigenvalue equation of the observable P(A), say, for the individual system S_i is written here as

$$P(A)^{(i)}\varphi_k^{(i)} = A_k^{(i)}\varphi_k^{(i)}, \quad k \in \{1,2\}, \quad A_1^{(i)} = 1, \quad A_2^{(i)} = 0 \,.$$

For an arbitrary state $\varphi \in H_S$ with $(\varphi,\varphi) = 1$ the product state $(\varphi)^\infty = \varphi^{(1)} \otimes \varphi^{(2)} \cdots$ with $\varphi^{(i)} \in H(S_i)$ is a state in $H_S^{(\infty)}$, i.e. $(\varphi)^\infty \in H_S^{(\infty)}$.

If one performs $P(A)$ – measurements on each system S_i one obtains a sequence $s_{\{l\}}$ of $P(A)$ – eigenvalues $A_{l(i)}^{(i)} \in \{0,1\}, l = \{l_1, l_2, \cdots\}$ and $l_\mathrm{i} \in \{1,2\}$. Let $\sum = \{s_{\{l\}}\}$ be the nondemumerable set of sequences of this kind. Any subset $\sum^{(\alpha)} \subseteq \sum$ describes a property. E.g. the set $\sum^{(p)}$ of sequences $s_{\{l\}}$ with the “probability **p** law of large number property” (with respect to $A^{(i)} = 1$) reads

$$\sum^{(p)} = \left\{ (s_{\{l\}}) : \lim_{N\to\infty} \frac{1}{N} \sum_1^N A_{l(i)}^{(\mathrm{i})} = p \right\}$$

For any subset we introduce an indicator function $F^{(\alpha)}(s_{\{l\}})$ by $F^{(\alpha)}(s_{\{l\}}) = 1$ if $s_{\{l\}} \in \sum^{(\alpha)}$ and $F^{(\alpha)}= 0$ otherwise.

According to the spectral theorem by an indicator function $F^{(\alpha)}$ a projection operator $P^{(\alpha)}$ is uniquely defined in $H_S^{(\infty)}$. Hence, for the product state $(\varphi)_l^\infty$ we obtain

$$P^{(\alpha)}(\varphi)_l^\infty = F^{(\alpha)}(s_{\{l\}})(\varphi)_l^\infty \,.$$

In the nonseparable Hilbert space $H_S^{(\infty)}$ there are many product states $(\psi)_k^\infty = \psi_{k(1)}^{(1)} \otimes \psi_{k(2)}^{(2)} \ldots$ which are not superpositions of the states $(\varphi)_l^\infty$. For these states one obtains

$$|P^{(\alpha)}(\psi)_k^\infty|^2 = \int F^{(\alpha)}(s_{\{l\}})d\mu \qquad (*)$$

where the measure μ depends on the state $(\psi)_k^\infty$ and on the measured observable.

[14] Gutman (1995, CPQ)
[15] von Neumann (1938, IDP)

For explicitly calculating the measure μ we consider first the case of arbitrary finite N. Using the spectral decomposition of the projection operator

$$P^{(\alpha)} = \sum_l F^{(\alpha)}(s_{\{l\}}) P[(\varphi)_l^N]$$

we obtain for $\psi^N = \psi_1 \otimes \cdots \otimes \psi_N$

$$P^{(\alpha)} \left|\psi^N\right\rangle = \sum_l F^{(\alpha)}(s_{\{l\}}) \langle l \left|\psi^N\right\rangle |l\rangle$$

with $|l\rangle := \left|(\varphi)_l^N\right\rangle$, and thus

$$\left|P^{(\alpha)} \left|\psi^N\right\rangle\right|^2 = \sum_l F^{(\alpha)}(s_{\{l\}} \left|\langle l|\psi^N\rangle\right|^2 .$$

The measure μ assigns probabilities to sequences. With the probabilities $p_j = |\langle\varphi_1^{(j)}|\psi_{l(j)}^{(j)}\rangle|^2$ for the values $A_{l(j)}^{(j)} = 1$ for N independent A-values the measure μ reads

$$\mu\left(s_{\{l\}} : A_{l(1)}^{(1)} = \alpha(1), \cdots A_{l(N)}^{(N)} = \alpha(N)\right)$$
$$= p_1^{\alpha(1)}(1-p_1)^{1-\alpha(1)} \cdots p_N^{\alpha(N)}(1-p_N)^{1-\alpha(N)} .$$

We are now in the position to discuss the main problem. Let $\mathrm{P}^{(1/2)}$ be the projection operator of the "probability p = ½ law of large number property" of a sequence $s_{\{l\}}$:

$$F^{(1/2)}(s_{\{l\}}) = 1 \text{ if } \lim_{N\to\infty} 1/N \sum A_{l(i)}^{(i)} = \frac{1}{2} \text{ and } F^{(1/2)}(s_{\{l\}}) = 0 \text{ otherwise} .$$

This situation is realised, for example, if all systems S_i are prepared in states

$$\psi^{(n)} = \frac{1}{2}(1+e^{i\pi/2})\varphi_A^{(n)} + \frac{1}{2}(1-e^{i\pi/2})\varphi_{\neg A}^{(n)}$$

and $P(A)^{(n)}$ is measured. The post-premeasurement mixed states read in this case

$$W_S^{(n)'} = \frac{1}{2}P[\varphi_A^{(n)}] + 1/2P[\varphi_{\neg A}^{(n)}]$$

and the probability p to find the A-value $A^{(n)} = 1$ is then given by $p = |\langle\varphi_1^{(n)}|\psi^{(n)}\rangle|^2 = 1/2$ for each n. The classical law of large numbers asserts in this case that the probability for the relative frequency 1/2 for the value $A = A_1 = 1$ in a sequence $s_{\{l\}}$ is equal to 1, i.e.

$$\int F^{(1/2)}(s_{\{l\}}) d\mu(s_{\{l\}}) = 1$$

This means that the relative frequency of $A = 1$ for almost all sequences amounts 1/2. Together with equation (*) derived above we get the relation $|P^{(1/2)}(\psi)^\infty|^2 = 1$ and thus

$$P^{(1/2)}(\psi)^\infty = 1 \cdot (\psi)^\infty.$$

According to the realistic interpretation of quantum mechanics this eigenvalue equation means that the compound system $S^{(\infty)}$ *possesses* the property given by $P^{(1/2)}$. Hence this property, or the relative frequency value 1/2 property, pertains to the system without any reference to probability.

It has been pointed out by Gutmann, l.c. that this way of reasoning can be applied to any "tail-property" with probability 1, e.g. to randomness. Within the context of the present problem this means, that in quantum mechanics the probabilistic way of speaking can be replaced by statements, which *do not* refer to probability. This result justifies the interpretation of the more technical results of the present Chapter. Whereas in classical probability theory probabilistic statements like $p = 1/2$ can never be reduced to statements which, for an arbitrary large ensemble, hold with certainty but only to statements which are almost true, in quantum mechanics probability statements can be replaced by propositions which hold without reference to probability.

Concluding Remarks

In a photon split beam experiment for an individual system it is objectively undetermined whether the photon has the property A or the counter property $\neg A$. However, in spite of the objective indeterminacy of each individual system, for a sufficiently large ensemble of photons we observe a strict law. The relative frequency of systems with property A is given by $p_1(\delta) = \cos^2 \delta/2$ where δ is the phase in the preparation state.

This statistical law proves to be a "law without law". Indeed, for an ensemble $S^{(N)}$ of N identically prepared systems in state φ, in the post-premeasurement state the relative frequency f_A^N of property A is in general not an objective property. However, in the limit $N \to \infty$ the relative frequency of A becomes an objective property of the compound system with the value $p(\varphi, A) = \cos^2 \delta/2$.

Formally this means that the probability for observing the relative A-frequency $\cos^2 \delta/2$ in a sequence of A-measurements is equal to 1. In terms of quantum mechanics this means that the "probability $p = \cos^2 \delta/2$ law of large numbers property" pertains to the system $S^{(\infty)}$ as an objective property. In this way the statistical laws of quantum mechanics emerge from the probability free theory and are – in Wheelers terminology – laws without laws.

13

Quantum Logic

Why are laws of nature valid? This general question in the title of Part III will be applied in this chapter to the laws of quantum logic. In this field we are confronted with a new situation and with new possible answers to the main question of this chapter. We will investigate here two questions, which, at first glance, seem to be alternative. Firstly we ask, whether the laws of quantum logic are genuine laws of nature that can be verified or falsified exclusively by experimental evidence. If a law of nature holds rigorously but only for empirical reasons then we will call it a *proper* law of nature. Proper laws of nature are contingent laws, which hold strictly but without logical necessity. Hence, our first question is whether the laws of quantum logic are proper laws of nature.

There are, however, also laws that hold not only rigorously in nature but also for a priori reasons. A law of this kind, which is not contingent but necessary will be called an *improper* law of nature. In the preceding chapters we found various improper laws of nature in logic, arithmetic, geometry, and statistics. In all these cases the laws hold with necessity, but the particular reasons for their validity are quite different. Hence, our second question will be whether the laws of quantum logic are improper laws of nature in the sense mentioned above. Quantum logic is often considered as a new logic that holds in the domain of quantum physics and must be used in this realm instead of the old classical logic. Hence we will ask whether quantum logic is a genuine logic which can be justified as other logical systems by purely a priori reasoning. If this were the case, then the laws of quantum logic would be improper laws of nature.

13.1 Are the Laws of Quantum Logic Laws of Nature?

13.1.1 Arguments Pro – What is Quantum Logic?

At first we will discuss the preliminary question what quantum logic is and we will give a preliminary answer. The final answer to this question will be given

at the end of this section, when the various arguments pro and contra have been discussed. In the first paper on this topic by Birkhoff and von Neumann[1] quantum logic is characterised as an atomic and orthomodular lattice, which fulfils the covering law. This view of the problem was further elaborated by Piron[2], Jauch[3] and others. According to these investigations quantum logic is a formal structure of quantum mechanics of an objects system S formulated in a Hilbert space H_S. Indeed, from quantum mechanics in Hilbert space we can obtain the lattice L_H of closed linear manifolds in that Hilbert space and this lattice turns out to be an atomic and orthomodular lattice L_Q^* which fulfils the covering law.

The name "quantum logic" comes from a formal analogy between the Hilbert lattice L_H of and the Boolean lattice L_C of classical logic. Moreover, the elements of the lattices L_H, the subspaces of Hilbert space correspond to projection operators, i.e. to observables with two values (eigenvalues) 0 and 1. Hence, these observables or properties are related to propositions with truth values 1 (yes) and 0 (no). If the property in question pertains to the quantum system, the corresponding proposition is said to be true, and if the counter property pertains to the system, the proposition is said to be false.

For more details let us consider a Hilbert space H_S and subspaces M_A, M_B,... with $M_A \subseteq H_S$, $M_B \subseteq H_S, \ldots$ etc. To every subspace and to every element $f \in H_S$ there exists a unique decomposition of f such that

$$f = f_A + f_{\neg A} \quad \text{with} \quad f_A \in M_A \quad \text{and} \quad f_{\neg A} \in M_{\neg A}$$

where $M_{\neg A}$ is the subspace, which is completely orthogonal to M_A, i.e.

$$M_{\neg A} := \{f \in H_S : (f, g) = 0\} \quad \text{for all} \quad g \in M_A.$$

The element f_A is called the projection of f onto M_A. The function $P_A := H_S \to M_A$ with $P_A f = f_A$ defines an operator P_A, the projection operator which maps an element $f \in H_S$ into its projection $f_A \in M_A$. The operator P_A has two eigenvalues 1 and 0 such that

$$P_A f = f \quad \text{if and only if } f \in M_A; \quad P_A f = 0 \text{ if and only if } f \in M_{\neg A}$$

Since the observable P_A has two eigenvalues we will call it also a "property" P(A).

For a given system S with the pure state $\varphi \in H_S$ the concept of an elementary proposition A can then be defined by

$$\begin{aligned} &\text{A is true} \Leftrightarrow \text{P(A) pertains to S} \Leftrightarrow \varphi \in M_A \\ &\text{A is false} \Leftrightarrow P(\overline{A}) \text{ pertains to S} \Leftrightarrow \varphi \in M_{\mathrm{A}}\ . \end{aligned}$$

1 Birkhoff, von Neumann (1936, LQM).
2 Piron (1964, AQu).
3 Jauch (1968. FQM).

Here we introduced the counter proposition $\overline{A}$ of A, where $\overline{A}$ is true iff A is false, and $\overline{A}$ is false iff A is true. An elementary proposition can be tested by measuring the observable P_A. Since the result of this measurement is either the value 1 or the value 0 the proposition A is either true or false. Elementary propositions of this kind are called to be *value definite.* They have always a well-defined truth value.

On the set $L_H = \{M_\lambda\}_\lambda$ of subspaces we will define now some operations and one relation which lead to an interesting algebraic structure.

(α) A binary relation $R \subseteq L_H \times L_H$ which is given by the set-theoretical inclusion "$\subseteq$" between subspaces;
(β) The one-place operation $\Theta_\neg : L_H \to L_H$ which is given by the completely orthogonal subspace, i.e. $\Theta_\neg(M_A) = M_{\neg A}$.
(γ) The two place operation $\Theta_\cap : L_H \times L_H \to L_H$ which is given by the intersection of two subspaces M_A and M_B, i.e. $\Theta_\cap(M_A, M_B) = M_A \cap M_B = \{f \in H_S\text{: } f \in M_A \text{ and } f \in M_B\}$.
(δ) The two-place operation $\Theta_\cup$: $L_H \times L_H \to L_H$ which is given by the span of two subspaces M_A and M_B, i.e. $\Theta_\cup(M_A, M_B) = M_A \cup M_B = \{f \in H_S\text{: } f = \alpha g + \beta h\}$ with $g \in M_A$, $h \in M_B$ and complex numbers α and β.

The structure which is induced on the set L_H of subspaces by the relation R and the three operations mentioned is the *Hilbert lattice* $L_H = \{L_H; \subseteq, \cap, \cup, \neg\}$. Apart from many other properties, L_H is an atomic orthomodular lattice, which fulfils the covering law. A lattice of this kind will be denoted here by L_Q^*.

Subspaces correspond to projection operators and projection operators correspond to value definite propositions. Hence, the lattice structure which is generated on the set of subspaces by the relation $\subseteq$ and the operations $\cap$, $\cup$, and $\neg$ induces a lattice structure on the set of propositions. The lattice of propositions is an orthomodular lattice L_Q with a zero element Λ and a unit element V. In addition, it is atomic and it fulfils the covering law. If these properties are included the lattice will be denoted by L_Q^*. Within the lattice of propositions we make use of a new notation and replace the relation $\subseteq$ by $\leq$ and the operations $\cap$ and $\cup$ by $\wedge$ and $\vee$, respectively. If we interpret "$\wedge$" by "and", "$\vee$" by "or", "$\neg$" by "not", and "$\leq$" by an implication, then the lattice L_Q^* shows many similarities with the well known Boolean lattice L_C of classical logic. This analogy between L_C and L_Q^* was the reason for many authors to call the lattice L_Q^* of quantum propositions "quantum logic"

The lattices L_Q and L_Q^* of propositions can be characterised by the following axioms:

$$
\begin{array}{ll}
L_Q(1) \qquad & A \leq A \\
& A \leq B, B \leq C \Rightarrow A \leq C \\
& A \leq B, B \leq A \Rightarrow A = B
\end{array}
$$

With respect to the implication "$\leq$" L_Q is a partially ordered set (poset). The third law defines the equivalence relation "$=$".

$$
\begin{array}{ll}
L_Q(2) & A \wedge B \leq A \\
& A \wedge B \leq B \\
& C \leq A, C \leq B \Rightarrow C \leq A \wedge B \\
L_Q(3) & A \leq A \vee B \\
& B \leq A \vee B \\
& A \leq C, B \leq C \Rightarrow A \vee B \leq C
\end{array}
$$

According to these axioms $L_Q(2)$ and $L_Q(3)$ $A \wedge B$ is the infimum and $A \vee B$ the supremum of A and B with respect to the implication relation.[4] Hence for any two elements A and B there exists an infimum and a supremum. A partially ordered set with this property is called a lattice.

$$
\begin{array}{lll}
L_Q(4) & \Lambda \leq A, A \leq V & \text{for all } A \in L_Q \\
& A \wedge \neg A \leq \Lambda & \\
& V \leq A \vee \neg A & \\
& A = \neg(\neg A) & \\
& A \leq B \Rightarrow \neg B \leq \neg A &
\end{array}
$$

If in a lattice with a zero element Λ and a unit element V an automorphism $A \to \neg A$ is defined which fulfils $L_Q(4)$ then this lattice is said to be *orthocomplemented*. An orthocomplemented lattice will be denoted by L_O. The propositions Λ and V are defined here as the smallest and largest propositions of L_O with respect to the implication relation. Hence Λ (*falsum*) is the false proposition and V (*verum*) the true proposition. Accordingly, by means of these special propositions one can express the truth and the falsity of a proposition A by

$$V \leq A \ (A \text{ is logically true}), \qquad A \leq \Lambda \ (A \text{ is logically false}) .$$

$$L_Q(5) \ B \leq A, C \leq \neg A \Rightarrow A \wedge (B \vee C) \leq B \ \text{(Orthomodularity)}$$

An orthocomplemented lattice which fulfils axiom $L_Q(5)$ is called *orthomodular*. It is denoted here by L_Q. In the lattice L_Q an element $\alpha \neq \Lambda$ is called an *atom* if for any $X \in L_Q, \Lambda \leq X \leq \alpha$ implies either $X = \Lambda$ or $X = \alpha$.

$L_Q(6)$ For any element $A \in L_Q$ there exists an atom α with $\alpha \leq A$.

A lattice which fulfils $L_Q(6)$ is called *atomic* (atomicity).

$L_Q(7)$ Let $\alpha \in L_Q$ be an atom. For all elements A and X of L_Q
$A \leq X \leq A \vee \alpha$ implies $X = A$ or $X = A \vee \alpha$ (covering law) .

A lattice L_Q that is atomic and fulfils the covering law will be denoted by L_Q^*.

[4] The physical meaning of the operations $A \wedge B$ and $A \vee B$ in particular for incommensurable propositions A and B will be explained in Sects. 13.1.2.2 and 13.2.1.1.

On the basis of these results and analogies the lattices L_Q and L_Q^* of quantum mechanical propositions could be considered as a purely empirical structure. With respect to the Hilbert space background of L_Q and L_H Putnam[5] says that we can "*read the logic off from the Hilbert space*". Hence, one gets the impression that the question in the title of Putnam's article "Is logic empirical?" should be answered in the positive sense, at least in the context of quantum logic. If this way of reasoning were the only way for justifying quantum logic, then quantum logic would indeed be a contingent formal structure of physics, which could be falsified or verified by experience. In this case quantum logic would be a proper law of nature. Hence our first and preliminary conclusion is that in spite of some formal similarities between the lattices L_Q^* and L_C the laws of quantum logic are laws of nature.

13.1.2 Objections Against 13.1.1

There are, however, important arguments against our first conclusion. The similarity between the lattices L_Q^* of quantum logic and the Boolean lattice L_C of classical logic is not only a vague formal analogy but a strong indication that also L_Q^* is a genuine logic. The relation between classical mechanics and classical logic can illustrate this conjecture and thus serve as a guiding principle for the interpretation of the abstract lattice L_Q^*. The detailed comparison of the lattices L_Q^* and L_C then shows that there is more agreement than disagreement and there are only small, though important differences between these two lattices.

13.1.2.1 The Logic of Classical Mechanics

In classical mechanics an object system S with n degrees of freedom is described by a phase space Γ_S with $2n$ coordinates $\{q_i, p_i\}$ which correspond to generalised position and momentum coordinates. The state of S at a time t_0 is given by a point $X(t_0) = X_0 = \{q_i^0, p_i^0\}$ of Γ_S. Observables are given by functions which map Γ_S into a real space R^N, e.g. the position $\mathbf{q}$ of a mass point is a function $\mathbf{q} : \Gamma_S \to R^3$. *Properties* are the simplest observables that take only two values 0 and 1, say. Properties are given by subspaces Γ_S^A, Γ_S^B, ... of Γ_S and will be denoted by $P(A), P(B), \ldots$, respectively. Hence, a property is a function $P(A)$: $\Gamma_S^A \to \{0, 1\}$ with $P(A)(X) = 1$ if $X \in \Gamma_S^A$ and $P(A)(X) = 0$ if $X \notin \Gamma_S^A$. We say that the observable take the value 1 or the property $P(A)$ pertains to S at time t_0 if $X_0 \in \Gamma_S^A$, and that $P(A)$ does not pertain to S if $X_0 \notin \Gamma_S^A$. In terms of propositions $A, B, \ldots$ this means that a proposition A is true if the property $P(A)$ pertains to S.

Similarly as in the case of quantum logic we are no longer interested in the space Γ_S but in the algebraic structure of the set $\{\Gamma_S^A\}$ of subsets of Γ_S, i.e.

[5] Putnam (1969, ILE).

in the structure of the set $P(\mathrm{S})$ of propositions about system S. With respect to the relation $\leq$, which is given by the set theoretical inclusion we have

$$\begin{array}{ll} L_C(1) & A \leq A \\ & A \leq B, B \leq C \Rightarrow A \leq C \\ & A \leq B, B \leq A \Rightarrow A = B \end{array}$$

i.e. $P(\mathrm{S})$ is a partially ordered set where the third law defines the equivalence of propositions. With respect to the two-place operations $A \wedge B$, given by the set theoretical intersection and $A \vee B$, given by the set theoretical union, we have

$$\begin{array}{ll} L_C(2) & A \wedge B \leq A \\ & A \wedge B \leq B \\ & C \leq A, C \leq B \Rightarrow C \leq A \wedge B \\ L_C(3) & A \leq A \vee B \\ & B \leq A \vee B \\ & A \leq C, B \leq C \Rightarrow A \vee B \leq C \end{array}$$

According to these axioms for any two elements $A, B \in P(S)$ with respect to the relation "$\leq$" there exists an infimum $A \wedge B$ and a supremum $A \vee B$. Hence $P(\mathrm{S})$ is a complete lattice denoted here by L_C. In L_C there exists a minimal element Λ which corresponds to the empty space $\Gamma_S^0 = \emptyset$ and a maximal element V which corresponds to the entire phase space Γ_S. Accordingly, for any $A \in L_C$ it holds $\Lambda \leq A \leq V$.

Furthermore, in a lattice L_C with elements Λ and V an automorphism $A \rightarrow \neg A$ is defined which fulfils the laws

$$\begin{array}{lll} L_C(4) & \Lambda \leq A,\ A \leq V & \text{for all } A \in L_C \\ & A \wedge \neg A \leq \Lambda & \\ & V \leq A \vee \neg A & \\ & A = \neg(\neg A) & \\ & A \leq B \Rightarrow \neg B \leq \neg A & \end{array}$$

For a proposition A which corresponds to a subspace $\Gamma_S^A \subseteq \Gamma_S$ the proposition $\neg A$ corresponds to the complementary subspace $\Gamma_S^{\neg A} = \Gamma_S \setminus \Gamma_S^A$. A lattice which fulfils the laws $L_C(4)$ is called orthocomplemented and the element $\neg A$ is said to be the orthocomplement to A.

Up to this point the lattice L_C given by $L_C(1\text{–}4)$ fulfils the same laws as the lattice L_Q of quantum mechanical propositions. The following law provides, however, remarkable differences between the two lattice structures. The lattice L_C of propositions is *distributive*, i.e. for any three elements $A, B, C \in L_C$ we have

$$L_C(5) \quad A \wedge (B \vee C) \leq (A \wedge B) \vee (A \wedge C)$$

that corresponds to the well known set theoretical relation $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. It is obvious that the distributive law $L_C(5)$ is stronger

than the corresponding orthomodular law $L_Q(5)$, since in L_Q^* distributivity implies $L_Q(5)$, but the inverse is not true. A more profound comparison of $L_Q(5)$ and $L_C(5)$ will be presented below.

In L_C a proposition $A \neq \Lambda$ is called an *atom*, if for any $X \in L_C, \Lambda \leq X \leq A$ implies $X = \Lambda$ or $X = A$. The atoms of L_C correspond to points in the phase space Γ_S. Hence we have

$L_C(6)$ For any $A \in L_C$ there exists an atom α with $\alpha \leq A$ (atomicity). In terms of phase space this means that any subspace contains at least one point. Furthermore we have the corollary

$L_C(7)$ Let α be an atom. For all propositions A, X of L_C
$$A \leq X \leq A \vee \alpha \text{ implies } X = A \text{ or } X = A \vee \alpha \text{ (covering law)} .$$

The complete, orthocomplemented, distributive and atomic lattice L_C is called a Boolean propositional lattice.

L_C is the lattice of classical propositional logic. Here we obtained this lattice by investigating the phase space of classical mechanics and by making use of a correspondence between mechanical yes–no propositions and subspaces of the phase space Γ_S of a system S. Hence, in the sense of Putnam[6] we could say that we can "read the classical logic off from the phase space". Clearly, this is correct but it does not imply that classical logic is empirical and thus a proper law of nature. It is well known that there are other strategies for establishing the propositional lattice L_C without explicit recourse to physical experience, which allow for a logical interpretation of the lattice L_C.[7] These arguments show that also in the case of quantum logic the derivation of the lattice L_Q^* from Hilbert space must not be considered as a sufficient reason for the statement that quantum logic is empirical and thus a proper law of nature.

13.1.2.2 Comparison of the Lattices L_Q^* and L_C

The Boolean lattice L_C admits a logical interpretation. Since the lattice L_Q^* has a very similar structure as the lattice L_C one could guess that also L_Q^* admits a logical interpretation which is, perhaps, slightly different from the interpretation of L_C. For the test of this conjecture we will briefly investigate some differences between L_C and L_Q^*.

In L_C we can define a new two-place operation $A \to B$, the *material implication*, by the laws
$$A \wedge (A \to B) \leq B \tag{1}$$
$$A \wedge X \leq B \Rightarrow X \leq A \to B \; . \tag{2}$$

Hence, $A \to B$ is the largest element which satisfies the *modus ponens* relation $A \wedge X \leq B$ and it is uniquely defined by (1) and (2) in L_C. The element $A \to B$ can be expressed by the other operations according to

[6] Putnam (1969, ILE).
[7] E.g. Lorenzen (1980, MeM).

$$A \to B = \neg A \vee B\ . \tag{3}$$

The denotation "material implication" is motivated by the fact that the proposition $A \to B$ is true if and only if the relation $A \leq B$ holds, i.e.

$$V \leq A \to B \Leftrightarrow A \leq B\ . \tag{4}$$

The existence of a material implication is often considered as an inevitable property of a lattice which allows for a logical interpretation since any logical inference makes use of the "modus ponens" law. In the orthomodular lattice L_Q^* an element $A \to B$ which fulfils (1) and (2) does not exist. This deficiency of a material implication in L_Q^* was often considered as a striking argument that L_Q^* cannot be interpreted as a logic.[8] In particular, without a material implication the *modus ponens* law $A \wedge (A \to B) \leq B$ cannot be formulated as an implication.

We can, however, overcome this problem in the following way. In L_Q^* we define an operation $A \to B$ by the two laws

$$A \wedge (A \to B) \leq B \tag{1*}$$

$$A \wedge X \leq B \Rightarrow \neg A \vee (A \wedge X) \leq A \to B \tag{2*}$$

This "material quasi implication" fulfils the modus ponens law (1*) and it is uniquely defined in L_Q^* by (1*) and (2*). In L_Q^* the element $A \to B$ can be expressed by the other operations as

$$A \to B = \neg A \vee (A \wedge B)\ . \tag{3*}$$

Furthermore, from (1*) and (2*) it follows

$$V \leq A \to B \Leftrightarrow A \leq B \tag{4*}$$

which means again that $A \to B$ is true iff $A \leq B$ holds.

It should be emphasised that the conditions (1*), (2*), and (3*) are relaxations of the conditions (1), (2), and (3) which are satisfied in a Boolean lattice L_C. In fact, in an orthocomplemented lattice L_O the conditions (1) and (2) imply the weaker conditions (1*) and (2*). In addition, in L_O the material implication $\neg A \vee B$ and the material quasi implication $\neg A \vee (A \wedge B)$ are in general only connected by the implication $\neg A \vee (A \wedge B) \leq \neg A \vee B$, whereas in a Boolean lattice L_C distributivity implies $\neg A \vee (A \wedge B) = \neg A \vee B$ and thus the elements agree. Hence, the material quasi implication seems to be a very convenient generalisation of the classical material implication, which fulfils the syntactical requirements for a logical interpretation of the lattice L_Q^*.

The difference between the lattices L_C and L_Q^* can be further illustrated by the following observation. In an orthocomplemented lattice L_O we can define a binary relation $K \subseteq L_O \times L_O$ called *commensurability* by

[8] Jauch, Piron (1970,WQL), pp. 173–174.

$$(A, B) \in K \Leftrightarrow A = (A \wedge B) \vee (A \wedge \neg B) \tag{5*}$$

and denote it by $A \sim B$. (Note, that in L_C this relation holds for any pair (A, B).) The name "commensurability" is motivated by the realisation of the lattice L_O by subspaces of a Hilbert space. In fact, the relation K holds for two subspaces M_A and M_B if and only if the corresponding projection operators P_A and P_B commute, i.e. the observables P_A and P_B are *commensurable* in the usual sense of quantum mechanics. Here, however, we are not concerned with subspaces and consider the relation K as a purely abstract relation defined by (5*).

From this definition it follows that in L_O the partial ordering relation R is contained in K, i.e. $R \subseteq K \subseteq L_O \times L_O$. Hence, for any pair (A, B) we have $A \leq B \Rightarrow A \sim B$. In L_O the relation K is, in general, not symmetric. The symmetry is rather a condition, which is fulfilled if and only if the lattice L_O is orthomodular. Hence, in an orthomodular lattice L_Q^* we have $A \sim B \Rightarrow B \sim A$. According to its physical meaning the relation of commensurability should be symmetric. Hence it is obvious that in the Hilbert lattice L_H the relation K is in fact symmetric. It is an interesting result that precisely the orthomodular law $L_Q(5)$ is necessary and sufficient for the symmetry of K in L_O.

The relation $A \sim B$ of commensurability can also be expressed by a "commensurability proposition"

$$k(A, B) := (A \wedge B) \vee (A \wedge \neg B) \vee (\neg A \wedge B) \vee (\neg A \wedge \neg B) \; .$$

which is true if and only if $A \sim B$ holds, i.e.

$$V \leq k(A, B) \Leftrightarrow A \sim B \; . \tag{6*}$$

Accordingly, the proposition "incommensurability" is given by $\overline{k}(A, B) = \neg k(A, B)$. On the basis of these definitions we can now answer the question raised in footnote 4. If A and B are incommensurable, then $A \wedge B$ is false and $A \vee B$ is true irrespective of the content of the propositions A and B, i.e.

$$V \leq \overline{k}(A, B) \Rightarrow A \wedge B \leq \Lambda \quad \text{and} \quad V \leq A \vee B \; . \tag{7*}$$

This result is further illustrated in Sect. 13.2.1.1 by means of the proof trees Fig. 13.3 and Fig. 13.4.

In order to further explain the meaning of the commensurability relation, we investigate the relation between commensurability and distributivity. The lattice L_Q^* is not distributive, i.e. the distributive law $A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$ does not hold generally (in the direction "$\leq$") in the orthomodular lattice L_Q^*. However, if the elements B and C are both commensurable with A, distributivity can be demonstrated, i.e. in L_Q^* we have

$$A \sim B, A \sim C \Rightarrow A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C) \; . \tag{8*}$$

This law of "weak distributivity" shows again that the orthomodular lattice L_Q^* is a relaxation of the orthocomplemented and distributive lattice L_C. The meaning of "weak distributivity" is further illustrated by the fact that the relation K of commensurability is closed with respect to the lattice operations $\wedge$, $\vee$, and $\neg$, i.e.

$$A \sim B, A \sim C \Rightarrow A \sim (B \wedge C), A \sim (B \vee C), A \sim \neg A \, . \tag{9*}$$

It follows from this closure property of the relation K together with the "weak distributivity" law that three elements $A, B, C \in L_Q^*$ which are pairwise commensurable will generate a sublattice $L(A, B, C) \subseteq L_Q^*$ which is orthocomplemented and distributive, i.e. a Boolean sublattice.

13.1.3 Preliminary Answer

These considerations show that the orthomodular lattice L_Q^* is a relaxation of the Boolean lattice L_C which preserves several properties of L_C which are indispensable for a logical interpretation, at least in a weakened version. There is a uniquely defined material quasi implication, which allows for logical inferences by means of the modus ponens law. The weak distributivity law shows that whenever propositions are mutually commensurable all the laws of classical logic are preserved in a Boolean sublattice of L_Q^*. On the basis of these results our preliminary answer to the question in the title of Chap. 13.1 "*are the laws of quantum logic laws of nature*" is, that the laws of quantum logic hold as laws of nature but that a logical interpretation of these laws and thus an "a priori" justification which is independent of experience is still possible and has not been excluded.

13.2 Are the Laws of Quantum Logic a priori Valid?

13.2.1 Arguments Pro – On the a priori Justification of Quantum Logic

In order to demonstrate the a priori validity of quantum logic we will show that within the framework of a scientific language the laws of quantum logic follow from the most general pragmatic preconditions of this language. Hence we will show that the propositions of this language are governed by the laws of quantum logic irrespective of any physical experience. This is the usual meaning of the term "a priori". Accordingly, our task will consist of two parts. Firstly, we have to constitute a formal language of physical propositions and secondly, we have to establish the formal propositional logic of this language and to show that this logic agrees with quantum logic.

13.2.1.1 The Formal Language of Quantum Physics

We consider a scientific language S_Q with elementary propositions $A = A(S)$ which state that a property $P(A)$ pertains to the object system S. Accordingly, the proof of the elementary proposition A consists in a measurement of property $P(A)$ with positive outcome. The general possibilities for quantum measurements allow for the assumption that after the measurement of $P(A)$ we obtain either a positive or negative result.[9,10] Hence, an elementary proposition A can either be proved (result A) or disproved (result $\overline{A}$), where $\overline{A}$ is the counter proposition, and are thus value-definite. Furthermore, if after a successful proof of A a new proof attempt for A is made, then one obtains again the result A, if the applied measurement process is repeatable. However, if after a successful proof of A another elementary proposition B is proved, then a new proof attempt for proposition A will in general not lead to the previous positive result. Hence, two propositions A and B are in general not simultaneously decidable. This is only the case if the corresponding properties $P(A)$ and $P(B)$ are "commensurable". In this case we will call also the propositions A and B "commensurable".

One could think that the restrictions of simultaneous measurability mentioned represent an empirical element that is incorporated here into the language S_Q. Indeed, these restrictions are *motivated* by the well-known discovery of incommensurable properties in quantum mechanics. However, for the constitution of the formal language we have not made use of new empirical results, but we have rather dispensed with the assumption, which is tacitly made in the language S_C of classical physics, that arbitrary pairs of propositions can simultaneously be tested by measurements. Hence, the pragmatic preconditions of proving or disproving propositions which we assumed here, are based on less empirical assumptions than the pragmatics used in a language S_C of classical physics. This means that we presuppose here a *quantum pragmatics* which is weaker than the pragmatics of classical language and classical logic. As to the semantics, an elementary proposition $A(S)$ will be called to be *true* if in a measurement process the property $P(A)$ was shown to pertain to the system S. A proposition is said to be *false* if the counter proposition $\overline{A}$ was shown to be true, i.e. if the property $P(\overline{A}\ )$ was shown to pertain to system S. Conversely, the truth of A implies that proposition $\overline{A}$ is false. Since the concepts of truth and falsity are based here on the results of measurement processes for the set S_q^e of elementary propositions we have here a *realistic semantics* in the spirit of Aristotle and Tarski.

Elementary propositions $A, B, \ldots$ are assumed here to be incommensurable in general, i.e. not simultaneously or jointly decidable. If proposition A, say, was shown to be true, then after a proof attempt of B and irrespective of the result (B or B), a new proof attempt of A will in general not lead to

[9] Busch et al. (1996, QTM).

[10] For a more general language without this assumption cf. Mittelstaedt (1978, QuL).

the previous result. Instead, this result is available after the B-test only if A and B are commensurable. In a sequence of proofs the results are only *restrictedly available,* where the restrictions are given by incommensurabilities. For the definition of the connectives the restricted availability is very important. Compound propositions, i.e. the connectives will be defined here by the possibilities to prove or to disprove the propositions in question. For the sequential propositions the temporal order of proofs counts whereas for logical propositions the order of proofs is irrelevant. The most simple connective, the "sequential conjunction" $A \sqcap B$ (A and then B) is defined by the following attack and defence schema,

connective	denotation	attacks	defences
$A \sqcap B$	"A and then B"	1. $A?$, 2.$B?$	1.$A!$, 2.$B!$

where $A?$ means the challenge to prove A, and $A!$ the successful proof. This attack and defence scheme can be illustrated most conveniently by a proof tree which is chronologically ordered.[11,12] The first branching point corresponds to the test of A at t_1, the second one corresponds to the B-test at t_2.

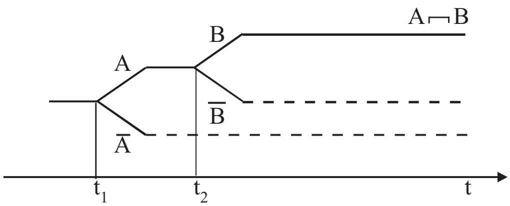


Fig. 13.1. Proof tree for the sequential conjunction $A \sqcap B$

The temporal order is fixed here, but the time difference $\delta t = t_2 - t_1 > 0$ may assume arbitrary positive values. There is one branch of success. In a similar way the one place operation "negation" may be introduced by the proof three for $\neg A$ (not A) with one branch for success and one branch for loss.

Obviously, the restricted availability of elementary propositions does not invalidate the definitions of the negation $\neg A$ and the sequential conjunction $A \sqcap B$. The negation is defined by one proof attempt and the sequential conjunction by two subsequent proof attempts. In both cases the restricted availability does not matter since repeated proof attempts do not occur here.

[11] Mittelstaedt (1978, QuL).
[12] Stachow (1980, LFQ).

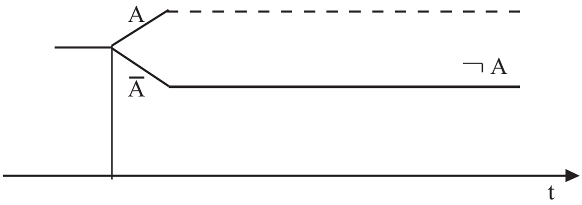


Fig. 13.2. Proof tree for the negation $\neg A$

However, the restrictions do matter if one tries to define the other connectives.[13,14]

The logical conjunction $A \wedge B$ is defined here by the following attack and defence scheme.

connective	denotation	attacks	defences
$A \wedge B$	A and B	$A?$, $B?$	$A!$, $B!$

Since the propositions are only *restrictedly available*, the proof tree for $A \wedge B$ consists of an infinite number of steps and cannot be reduced to two steps. If, however, A and B were commensurable, then it would be possible to reduce the proof tree to one A-proof and one B-proof. In order to achieve generally at a finite proof tree we make use of the commensurability proposition $k(A, B)$ which is defined to be true if and only if A and B are commensurable.The counter proposition is denoted here by $\overline{k}(A, B)$.[15] The logical conjunction $A \wedge B$ is then true if in addition to A and B also $k(A, B)$ is shown to be true. Hence, we have a proof tree with three subsequent tests at time values t_1, t_2, t_3. Since the conjunction $A \wedge B$ is understood as a simultaneous connective, the time differences $t_3 - t_2$ and $t_2 - t_1$ must be sufficiently small (Fig. 13.3).

The commensurability propositions $k(A, B)$ and $\overline{k}(A, B)$ are contingent propositions whose truth must be shown by a convenient sequence of measurements. We will not go into detail here. By means of the commensurability propositions $k(A, B)$ and $\overline{k}(A, B)$ one can define also the logical disjunction $A \vee B$ and the material implication $A \rightarrow B$ by proof trees with a finite number of steps. (Fig. 13.4) It should be emphasised, that we make use here of a semantics of truth which consists of two parts, the *realistic* semantics of elementary propositions and the *proof tree* semantics of the compound propositions. In order to incorporate both the measurements process and the attack and defence process into a unified semantical concept, we speak of a *process semantics.*

[13] Mittelstaedt, Stachow (1978, PEM).
[14] Stachow (1980, LFQ).
[15] Mittelstaedt (1978, QuL), (1986, SRP).

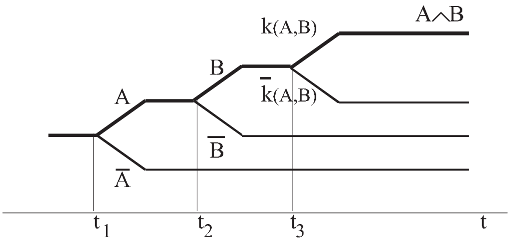


Fig. 13.3. Proof tree for the logical conjunction

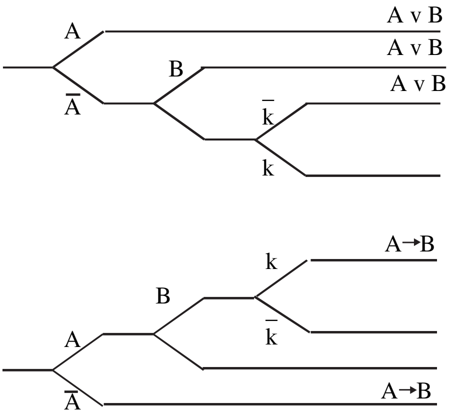


Fig. 13.4. Proof trees for the logical disjunction $A \vee B$ and the material implication $A \rightarrow B$

Furthermore, we can define here binary *relations* between propositions. The *proof equivalence* $A \equiv B$ means that A can be replaced in any proof tree by B without thereby changing the result of the proof tree. The binary relation of *value equivalence* $A = B$ means that A is true if and only if B is true.[16] The relation of *implication* $A \leq B$ can then be defined by $A \equiv A \wedge B$. Finally, we mention that $A \rightarrow B$ is true if and only if $A \leq B$ holds and that the commensurability proposition is true if and only if $A \leq (A \wedge B) \vee (A \wedge \neg B)$ holds.

The full language S_Q of quantum physics can then inductively be defined by the set $S_Q^{(e)}$ of elementary propositions, the commensurability propositions k and $\overline{k}$ and the connectives mentioned. Together with the relations "$\equiv$", "$=$", and "$\leq$" the quantum language S_Q reads

$$S_Q = \{S_Q^{(e)}; k, \overline{k}; \sqcap, \wedge, \vee, \rightarrow, \neg; \equiv, =, \leq \}$$

[16] If two propositions are *proof equivalent*, then they are also *value equivalent*. The inverse is not generally true. However, in classical language the two equivalence relations coincide.

13.2.1.2 Quantum Logic

Quantum logic is the formal logic of the quantum language S_Q and its syntax. The term "formal logic" is understood here as the totality of all formally true propositions. A proposition A is called "formally true" if it is true in the sense of the process semantics irrespective of the truth or falsity of the elementary propositions contained in the compound proposition A. It turns out that in quantum language S_Q there are less formally true propositions than in classical language S_C. In order to make these preliminary remarks more precise we will express the totality of all formally true propositions of quantum language by a calculus, the calculus L_Q of quantum logic.

There are, first of all, many formally true propositions of classical language S_C that are also formally true in quantum language S_Q. The value definiteness of elementary propositions implies that also all finitely connected propositions are value definite,[17] i.e. the proposition $A \vee \neg A$, the *tertium non datur law*, is formally true. The precondition that measurements are repeatable in principle implies that $k(A, A)$ is always true and hence $A \rightarrow A$, the *law of identity*, is formally true. In a similar way, it follows that $\neg(A \wedge \neg A)$, the *law of contradiction*, is formally true in quantum logic. The three cases mentioned are not very surprising since these formally true propositions contain only one proposition A. Hence, commensurability problems cannot appear. There are, however, also formally true propositions in quantum logic that contain two or more elementary propositions, where nothing is presupposed about their mutual commensurability. An example of this kind is the proposition $(A \wedge (A \rightarrow B)) \rightarrow B$, the *modus ponens law*, which is formally true in quantum logic irrespective of the truth or falsity of the commensurability proposition $k(A, B)$.

More important for the characterisation of quantum logic are those propositions which are formally true in classical logic but not in quantum logic. The shortest and in addition most important proposition which is formally true in classical logic but not in quantum logic is the proposition $A \rightarrow (B \rightarrow A)$. In classical logic, i.e. under the assumption of unrestricted availability of all propositions the proof tree for $A \rightarrow (B \rightarrow$A) contains only branches of success.

In quantum logic the situation is more complicated since the proof tree contains also the test of commensurability propositions $k(A, B)$. Only if the commensurability of A and B were presupposed, then the proof tree would contain only successful branches. This means that in general the proposition $A \rightarrow (B \rightarrow A)$ is not true and thus not formally true.

The totality of all propositions which are formally true even under the restrictions that are provided by the commensurability tests is called quantum logic. There are – as in classical logic – infinitely many propositions that are formally true in the sense of quantum logic. They can be summarised in a quantum logical calculus L_Q, which contains "beginnings" $\Rightarrow A \leq B$ and

[17] Mittelstaedt, Stachow (1978, PEM).

"rules" of the form $A \leq B \Rightarrow C \leq D$. For the formulation of this calculus we make again use of the two special propositions V (verum) and Λ (falsum) such that for all propositions $A \in S_Q$ the relations $\Lambda \leq A \leq V$ hold. If $A \to (B \to A)$ is true then the relation $A \leq (B \to A)$ holds. $A \leq B \to A$ implies B$\leq A \to B$ and vice versa and $A \leq B \to A$ holds if and only if $k(A, B)$ is true. Hence, in a calculus of quantum logic the commensurability propositions $k(A, B)$ can be eliminated by this implication and will no longer appear in its final formulation. The calculus L_Q of quantum logic reads:

The calculus of quantum logic	
1.1.	$\Rightarrow A \leq A$
1.2.	$A \leq B; B \leq C \Rightarrow A \leq C$
2.1	$\Rightarrow A \wedge B \leq A$
2.2.	$\Rightarrow A \wedge B \leq B$
2.3.	$C \leq A; C \leq B \Rightarrow C \leq A \wedge B$
3.1.	$\Rightarrow A \leq A \vee B$
3.2.	$\Rightarrow B \leq A \vee B$
3.3.	$A \leq C; B \leq C \Rightarrow A \vee B \leq C$
4.1.	$\Rightarrow A \wedge (A \to B) \leq B$
4.2.	$A \wedge C \leq B \Rightarrow A \to C \leq A \to B$
4.3.	$A \leq B \to A \Rightarrow B \leq A \to B$
4.4.	$B \leq A \to B; C \leq A \to C \Rightarrow B * C \leq A \to B * C$ $* \in \{\wedge, \vee, \to\}$
5.0	$\Rightarrow \Lambda \leq A, \Rightarrow A \leq V$
5.1.	$\Rightarrow A \wedge \neg A \leq \Lambda$
5.2.	$A \wedge C \leq \Lambda \Rightarrow A \to C \leq \neg A$
5.3.	$A \leq B \to A \Rightarrow \neg A \leq B \to \neg A$
5.4.	$\Rightarrow V \leq A \vee \neg A$

The Lindenbaum–Tarski algebra of the calculus L_Q is given by a complete orthomodular lattice L_Q. Subsets of mutual commensurable propositions constitute Boolean sublattices $L_B^{(i)} \subseteq L_Q$ of the lattice L_Q.[18] Moreover, if the entire quantum language S_Q refers to one quantum system S, then the lattice L_Q is atomic and fulfils the covering law.[19] Hence we arrive at the lattice L_Q^* which we obtained from Hilbert space in Sect. (13.1.1). This means that we have reconstructed this lattice operationally and shown in this way that L_Q^* allows for a logical interpretation.

These results show that quantum logic, which is represented by a lattice L_Q^*, is not only an empirical structure which – in the sense of Putnam – can be "read off from Hilbert space" but also a priori valid in the usual sense of this term. Indeed, quantum logic follows from the general pragmatic preconditions of a formal language of quantum physics independent of any empirical result.

[18] Birkhoff, von Neumann (1936, LQM).
[19] Stachow (1984, QLI).

Moreover, comparing quantum logic and classical logic shows that all propositions that are formally true in quantum logic are also formally true in classical logic, but the inverse is not true. There are infinitely many formally true propositions in classical logic that are not formally true in quantum logic. The reason for this important observation is, that quantum logic is based on the pragmatic precondition of the *restricted availability* of quantum propositions which is weaker and more restrictive than the corresponding precondition of *unrestricted availability* of all propositions of classical logic. Hence, compared with classical logic quantum logic is the weaker structure but for this reason also less dependent on empirical premises than classical logic.

13.2.2 Objections Against 13.2.1

The formal system of quantum logic which is represented by the orthomodular lattice L_Q^* was reconstructed in 13.2.1 as the formal logic of the language S_Q of quantum propositions. The formally true propositions of this logic are true in the sense of the used process semantics irrespective of the truth or falsity of the elementary propositions contained in it. Hence, the laws of quantum logic hold for all propositions of S_Q without any recourse to physical experience. For this reason the laws of quantum logic were considered as a priori valid.

However, the laws of quantum logic are dependent on the pragmatic preconditions of the language S_Q and hence on the underlying ontology O_Q. This dependence becomes obvious if one compares quantum logic with classical logic and quantum pragmatics with classical pragmatics, respectively. Since the pragmatic preconditions on the languages S_Q and S_C are based on the respective ontologies O_Q and O_C the formal logic of the languages S_Q and S_C will finally depend on the ontologies O_Q and O_C, respectively. Hence we should formulate our result in (13.2.1) more precisely and say that the laws of quantum logic hold a priori with respect to the ontological preconditions of quantum language, i.e. with respect to the quantum ontology O_Q.

On the basis of these arguments it becomes obvious that the a priori validity of quantum logic does not mean independence of *any* experience. Quantum logic is independent of the special experimental information that is contained in the elementary propositions. However, it is dependent on the most general experience that constitutes the quantum ontology O_Q. We mention here in particular the pragmatic precondition of the "restricted availability" which turned out to be the reason for the difference between quantum logic and classical logic.

The comparison between quantum logic and classical logic leads us to the final and most important argument. The laws of quantum logic hold a priori with respect to the quantum pragmatics and quantum ontology O_Q. Whenever quantum ontology is presupposed, quantum logic is valid for arbitrary propositions of the language S_Q. (The laws of classical logic hold in the same sense a priori with respect to classical pragmatics and classical ontology

O_C.) As mentioned above classical and quantum ontology are not alternative ontologies, which belong to different worlds, but quantum ontology O_Q is merely a relaxation of the stronger classical ontology O_C. For this reason, quantum logic is weaker than classical logic but more general. It holds also in the domain of classical ontology. Consequently, quantum logic depends on less empirical presuppositions than classical logic and is thus valid in a wider region of application than classical logic. It is, however, not unrestrictedly valid.

13.2.3 Answer to Question 13.2

Are the laws of quantum logic a priori valid? The operational reconstruction of the orthomodular lattice of quantum logic leads to a twofold answer to this question. *Firstly*, the laws of quantum logic are a priori valid in the same sense as the laws of classical logic with the remarkable difference, that quantum logic is even "more a priori" than classical logic since it depends on weaker ontological preconditions. *Secondly*, the apriority of quantum logic refers only to possible substitutions of truth values of elementary propositions – but not to the experience contained in the ontology. Quantum logic as well as classical logic depend on empirical ontological premises.

13.3 Concluding Answer to Question 13.1

Are the laws of quantum logic laws of nature? On the one hand, we could show (in 13.1.) that the lattice L_Q^* of quantum logic is an abstract structure of quantum mechanics in Hilbert space. Hence, the laws of quantum logic can be confirmed and shown to be satisfied by experiments and are – in this weak sense – laws of nature. On the other hand, we could also show (in 13.2.) that the laws of quantum logic follow from the pragmatic preconditions of the language of quantum physics and are, for that reason, a priori valid. Hence, they must not be considered as proper laws of nature. However, the apriority of quantum logic is somewhat invalidated by the fact that the pragmatic preconditions (value definiteness, repeatability, restricted availability etc.) and the underlying ontology depend – as in case of classical logic – on experience. For that reason the laws of quantum logic are a priori valid and thus not genuine laws of nature but with a small empirical impurity, which makes them – in a weak sense – being laws of nature.

References

Abbott, L. (1988, RKK) Das Rätsel der kosmologischen Konstanten, Spektrum der Wissenschaft, Juli 1988, pp. 92–99.

Ackermann, W. (1956, BSI) Begründung einer strengen Implikation, The Journal of Symbolic Logic, **21**, pp. 113–128.

Anderson, A.R., Belnap, N.D. (1975, Ent) Entailment. The Logic of Relevance and Necessity, Vol. 1, Princeton University Press, Princeton.

Anderson, J.L. (1967, PRP) Principles of Relativity Physics, New York, Academic Press.

Angelopoulos, A. et al. (1998, CPL) CPLEAR collaboration. In: Phys. Lett. B 444, p. 43f.

Aristotle (1985, CWA) The Complete Works of Aristotle. The Revised Oxford Translation, J. Barnes (ed.), Princeton UP, Princeton.

Aristotle (Met) Metaphysics. In: The Complete Works of Aristotle. The Revised Oxford Translation, Vol. 2, Barnes, J. (ed.), Princeton, 1985.

Aristotle (Phys) Physics. In: The Complete Works of Aristotle. The Revised Oxford Translation, Vol. 1, Barnes, J. (ed.), Princeton, 1985.

Aristotle (Cat) Categories. In: The Complete Works of Aristotle. The Revised Oxford Translation, Vol. 1, Barnes, J. (ed.), Princeton, 1985.

Aristotle (Heav) On the Heavens, in: The Complete Works of Aristotle, The Revised Oxford Translation, Vol. 1, Barnes J. (ed.), Princeton, 1985.

Arnold, V.I. (1963, SDP) Small Denominators II. Proof of a Theorem of A.N. Kolmogorov on the Preservation of Conditionally-Periodic Motions under a Small Perturbation of the Hamiltonian, Russ. Math. Surveys **18**, pp. 5 ff.

Aspect et al. (1982, ETB) Experimental Tests of Bell's Inequalities using Time-varying Analysers. Phys. Rev. Lett. **49**, pp. 1804–1807.

Augustin (Conf) Confessiones. Transl. F.J. Sheed, London 1943.

Barbour, J.B. (1989, ARM) Absolute or Relative Motion? A Study From a Machian Point of View of the Discovery and the Structure of Dynamical Theories. Cambridge University Press, Cambridge.

Barbour, J.B. (2001, GCB) On General Covariance and Best Matching. In: Callender, C. and Huggett, N. (2001, PMP) pp. 199–212.

Barrow, J., Tipler, F. (1986, ACP) The Anthropic Cosmological Principle, Oxford University Press, Oxford.

Barrow, J. (1988, WwW) The World within the World, Oxford University Press, Oxford.

Barrow, J. (1988, TOE) Theories of Everything, Oxford University Press, Oxford.

Barrow, J. (2002, CNt) Constants of Nature, Vintage, London.

Bauer, H. (1978, WGH) Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie, de Gruyter, Berlin.

Beltrametti, E.G., Maczynski, M.J. (1991, CNP) On a Characterization of Classical and Nonclassical Probabilities, J. Math. Phys, **32**, pp. 1280–6.

Bencivenga, E. (1986, FLg) Free Logics, in: Gabbay, D., Guenthner, F. (eds.) Handbook of Philosophical Logic, Vol. III, Reidel Publ. Co., Dordrecht, pp. 373–426.

Benedicks, M., Carleson, L. (1991, DHM) The Dynamics of the Hénon Map. Annals of Mathematics **133,** pp. 73–169.

Berry, M.V. et al. (eds.) (1987, DCh) Dynamical Chaos, Proc. Roy. Soc. London, **A. 413,** pp. 1844 ff.

Berry, M.V. (1989, PCG) Principles of Cosmology and Gravitation. Inst. of Physics Publishing, Bristol.

Birkhoff, G., v. Neumann, J. (1936, LQM) The Logic of Quantum Mechanics, Annals of Mathematics **37**, pp. 823–843.

Bohm, D. (1957, CCM) Causality and Chance in Modern Physics, Routledge & Kegan Paul, London.

Bohm, D. (1980, WIO) Wholeness and the Implicate Order, Routledge & Kegan Paul, London.

Boltzmann, L. (1868, SGL) Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten. Sitzungsber. der Kaiserlichen Akad. Wiss. (Wien) **58**, pp. 517–560. Reprinted in: Boltzmann, L. (1968, WAB), Vol. I, §5.

Boltzmann, L. (1896, EWB) Entgegnung auf die wärmetheoretischen Betrachtungen des Herrn E. Zermelo. In: Boltzmann (1968, WAb), Vol. III, §119.

Boltzmann, L. (1897, AGV) Über einige meiner weniger bekannten Abhandlungen über Gastheorie und deren Verhältnis zu derselben. In: Boltzmann (1968, WAb), Vol. III, §123.

Boltzmann, L. (1897, MSP) Über einen mechanischen Satz Poincaré's. In: Boltzmann (1968, WAb), Vol. III, §121.

Boltzmann, L. (1897, ZAM) Zu Herrn Zermelos Abhandlung, 'Über die mechanische Erklärung irreversibler Vorgänge'. In: Boltzmann (1968, WAb), Vol. III, §120.

Boltzmann, L. (1968, WAb) Wissenschaftliche Abhandlungen. 3 Volumes, Chelsea Publ. Co., New York.

Bolza, O. (1909, VVR) Vorlesungen über Variationsrechnung. Leipzig.

Bolzano, B. (1929, WSL) Wissenschaftslehre, 4 Vols. A. Höfer, W. Schultz, (eds.) F. Meiner Verlag, Leipzig (1929–31). Parts translated and edited by George, R., University of California Press, Berkeley 1972.

Bonk, S. (1998, KIA) Kausalität, Induktion und Außenwelt: David Humes skeptisch-naturalistische Epistemologie, Philosophia Naturalis, **35**, pp. 281–308

Bonola, R. (1955, NEG) Non-Euclidean Geometry, (including some papers by Lobachevsky and Bolyai) Dover, New York.

Boole, G. (1854, LoT) The Laws of Thought, Dover Edition, New York 1958.

Boole, G. (1862, ToP) On the Theory of Probabilities, Phil. Trans. Roy. Soc. (London) **152**, pp. 225–52. Reprinted in: Boole, G. (1952, SLP), pp. 386–424.

Boole, G. (1952, SLP) Studies in Logic and Probability. In: Collected Logical Works, Vol. I, ed. R. Rhees, Open Court Publ. Co., La Salle.

Boolos, G. (1998, LLL) Logic, Logic, and Logic. Harvard University Press, Cambridge, Mass.

Borsuk, K., Szmielew, W. (1960, FdG) Foundations of Geometry. Euclidean and Bolyai–Lobachewskian Geometry, Projective Geometry, North Holland, Amsterdam.

Brans, C., Dicke, R.H. (1961, MPR) Mach's Principle and a Relativistic Theory of Gravitation, Phys. Rev. **124**, pp. 925–935.

Breuer, Th. (1995, IAS) The Impossibility of Accurate State Self-Measurements. In: Philosophy of Science **62**, pp. 197–214.

Breuer, Th. (1996, SDQ) Subjective Decoherence in Quantum Measurements. In: Synthese **107**, pp. 1–17.

Breuer, Th. (1997, IOP) Ignorance of the Own Past. In: Erkenntnis **48**, pp. 1–7.

Breuer, Th. (1997, QMG) Quantenmechanik: Ein Fall für Gödel? Spektrum Verlag, Heidelberg.

Breuer, Th. (2001, UCP) Universal Causation and Predictability, in: Spohn, W. et al. (2001, CIC), pp. 163–73.

Brudno, A.A. (1983, ECT) Entropy and the Complexity of the Trajectories of a Dynamical System, Transactions of the Moscow Mathematical Society **2**, pp. 127–151.

Brush, S.B. (1985, KMH) The Kind of Motion we call Heat, North Holland, Amsterdam.

Bunge, M. (1967, SRI) Scientific Research. Vol. I, Springer, Berlin.

Bunge, M. (1967, SRII) Scientific Research, Vol. II, Springer, Berlin.

Bunge, M. (1967, FPh) Foundations of Physics. Springer, Heidelberg.

Bunge, M. (ed.) (1967, DSF) Delaware Seminar in the Foundations of Physics. Springer, Berlin.

Bunge, M. (1972, TAT) Time Asymmetry, Time Reversal, and Irreversibility. In: J.T. Fraser, F.C. Haber, G.H. Müller (eds.): The Study of Time. Springer, Berlin, pp. 122–130.

Bunge, M. (1973, MMM) Method, Model, and Matter. Reidel, Dordrecht.

Burgess, J.P. (1984, BTL) Basic Tense Logic. In: Gabbay, D., Guenther, F. (1984, HBP) pp. 89–133.

Busch, P., Mittelstaedt, P. (1991, POQ) The Problem of Objectification in Quantum Mechanics, Foundations of Physics **21**, pp. 889–904.

Busch, P. et al. (1995, OQP) Operational Quantum Physics, Springer, Heidelberg.

Busch, P. et al. (1996, QTM) The Quantum Theory of Measurement, 2nd. ed. Springer, Berlin, Heidelberg 1996.

Busch, P., Lahti, P. and Mittelstaedt, P. (1992,WOB) Weak Objectification, Joint Probabilities and Bell Inequalities in Quantum Mechanics, Found. of Physics, **22**, pp. 949–962.

Callender, C. and Huggett, N. (2001, PMP) Physics Meets Philosophy at the Planck Scale. Cambridge University Press, Cambridge.

Carmeli, M. (1982, CFG) Classical Fields: General Relativity and Gauge Theory. Wiley, New York.

Cartwright, N. (1983, LPL) How the Laws of Physics Lie, Clarendon Press, Oxford.

Cartwright, N. (1989, NCM) Nature's Capacities and their Measurement, Oxford University Press, Oxford.

Casati, G. and Chirikov, B.V. (eds.) (1994, QCh) Quantum Chaos Between Order and Disorder, Cambridge University Press, Cambridge.

Castellani, E., Mittelstaedt, P. (2000, LPP) Leibniz's Principle, Physics, and the Language of Physics, Found. of Physics **30**, pp. 1587–1604.

Chalmers, A.F. (1970, CPr) Curie's Principle, British Journal for the Philosophy of Science **21**, pp. 133–148.

Chellas, B.F. (1975, BCL) Basic Conditional Logic, Journal of Philosophical Logic **4**, pp. 133–153.

Chiao et.al. (1997, NGD) Negative Group Delay and "Fronts" in a Causal System, Physics Letters A **230**, pp. 133–138.

Chirikov, B.V. (1987, PCA) Particle Confinement and Adiabatic Invariance. In: Berry M.V. et al. (eds.) (1987, DCh), pp. 145–156.

Chirikov, B.V. (1991, TDQ) Time Dependent Quantum Systems, in: Giannoni, M.J. et al. (eds.) Chaos and Quantum Physics, Elsevier Publ.

Chirikov, B.V. (1991, PCh) Patterns in Chaos. In: Chaos, Solitons and Fractals 1(1), pp. 79–103.

Chirikov, B.V. (1992, LCh) Linear Chaos, in: Springer Proceedings in Physics 67, pp. 3–13.

Chirikov, B. (1996, NLH) Natural Laws and Human Prediction. In: P. Weingartner and G. Schurz (1996, LPL), pp. 10–33.

Cornfeld, Fomin, Sinai (1982, ETh) Ergodic Theory, Springer, New York.

Courant, R., Hilbert, D. (1968, MMP) Methoden der mathematischen Physik I, Springer, Heidelberg.

Curie, P. (1894, SPP) Sur la symmétrie dans les phénomènes physiques, symmetrie d'un champ électrique et d'un champ magnétique, Journal de Physique **3**, pp. 393–415.

DaCosta, N. (1974, TIF) On the Theory of Inconsistent Formal Systems, Notre Dame Journal of Formal Logic **15**, pp. 497–510.

Damour, Th. et al. (1988, LVG) Limits on the Variability of G Using Binary Pulsar Data, Phys. Rev. Lett. **61**, pp. 1151–54.

Damour, Th., Dyson F. (1996, OBT) The Oklobound on the Time Variation of the Fine Structure Constant Revisited, Nuclear Physics B **480**, pp. 37–54.

Davies, P.C.W. (1994, SuT) Stirring up Trouble. In: Halliwell et al. (1994, POT), pp. 119–130.

DeWitt, B.S. (1971, MUI) The Many-Universes Interpretation of Quantum Mechanics, in: Foundations of Quantum Mechanics, IL Corso, B. d' Espagnat, ed., Academic Press, New York, pp. 167–218.

Dirac, P.A.M. (1937, ROQ) The Reversal Operator in Quantum Mechanics, Bulletin de l'Académie des Sciences de l'URSS 1937, pp. 569–575.

Dirac, P.A.M., (1937, CCs) Cosmological Constants, Nature **139**, pp. 323–324.

Dirac, P.A.M., (1938, NBC) A New Basis for Cosmology. In: Proc. Royal. Soc., London, A **165**, pp. 199–208.

Dirac, P.A.M. (1973, FCD) Fundamental Constants and their Development in Time. In: Mehra, J. (ed.) (1973, PCN), pp. 45–54.

Duhem (1913, SdM) Le système du monde. Vol. I, Paris.

Dunn, J.M. (1986, RLE) Relevance Logic and Entailment, in: Gabbay, D.M., Guenthner, F. (eds.), Handbook of Philosophical Logic, Vol. III, Reidel, Dordrecht, pp. 117–224.

Dyson, F. (1972, FCT) Fundamental Constants and their Time Variation. In: Salam, A., Wigner, E.P. (1972, AQT), pp. 213–236.

Dyson, F. (1978, VCs) Variations of Constants. In: Lanutti, Williams (eds.) Current Trends in the Theory of Fields, American Institute of Physics.

Dziobek (1888, MTP) Die mathematischen Theorien der Planeten-Bewegungen. Leipzig, J.A. Barth.

Earman, J. (1986, PDt) A Primer on Determinism. Reidel, Dordrecht.

Eddington, A. (1928, NPW) The Nature of the Physical World. Cambridge University Press, Cambridge.

Eddington, A. (1939, PPS) The Philosophy of Science, Cambridge University Press, Cambridge.

Einstein, A. (1905, EBK) Zur Elektrodynamik bewegter Körper. In: Annalen der Physik **17** (1905) pp. 891–921. Engl. Transl. in: Lorentz et al. (1923, PRT) pp. 37–65.

Einstein, A. (1905, TKE) Ist die Trägheit eines Körpers von seinem Energiegehalt abhängig? In: Annalen der Physik **18**, pp. 639–641.

Einstein, A. (1908, RPF) Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen. In: Jahrbuch der Radioaktivität und Elektronik **4** (1907) pp. 411–462; **5** (1908) pp. 98–99.

Einstein, A. (1911, ESA) Über den Einfluss der Schwerkraft auf die Ausbreitung des Lichtes. In: Annalen der Physik **35**(1911) pp. 898–908. Engl. Transl. in: Lorentz et al. (1923, PRT) pp. 99–108.

Einstein, A. (1916, GAR) Die Grundlagen der allgemeinen Relativitätstheorie. In: Annalen der Physik **49** (1916) pp. 769–822. Engl. Transl. in : Lorentz et al. (1923, PRT) pp. 109–164.

Einstein, A. (1924, ÜdÄ) Über den Äther. Verhandlungen der Schweizerischen Naturforschenden Gesellschaft **105** (1924) pp. 85–93.

Einstein, A. Podolsky, B. and N. Rosen (1935, CQD) Can the Quantum-Mechanical Description of Physical Reality be Considered Complete, Phys. Rev. **47,** pp. 777–780.

Einstein, A. (1944, BRE) Bemerkungen zu Bertrand Russells Erkenntnistheorie. In: Russell (1944, PBR), pp. 278–290.

Einstein, A. (1949, EPS) Albert Einstein, Philosopher–Scientist, ed. P.A. Schilpp, Tudor, New York.

Ehlers, J. (1991, NLG) The Newtonian Limit of General Relativity. In: Ferrase, G. (ed.) Classical Mechanics and Relativity: Relationship and Consistency. Bibliopolis, Napoli, 1991, pp. 95–106.

Esfeld, M. (1999, HdQ) Der Holismus der Quantenphysik, Philosophia Naturalis **36**, pp. 157–181.

Essler, W. (1969, ELg) Einführung in die Logik, Kröner, Stuttgart.

Fahr, H.J. (1997, WKS) Zum Wachstum kosmischer Strukturen; Aspekte und Ideen zur Strukturierung des Universums. In: Grenzfragen 1997: Wachstum als Problem; Modelle und Regulation. Alber, Freiburg, pp. 13–51.

Farmer, J.D. (1982, IDP) Information Dimension and the Probabilistic Structure of Chaos. Zeitschrift für Naturforschung **37a,** p. 1304.

Farmer, J.D. (1985, SDP) Sensitive Dependence on Parameters in Nonlinear Dynamics, Phys. Rev. Lett. **55**, pp. 351–354.

Feferman, S. (1964, SPA) Systems of Predicative Analysis, Journal of Symbolic Logic **29**, pp. 1–30.

Feynman, R.P. (1964, LPh) Lectures on Physics. Addison-Wesley, London.

Feynman, R.P. (1967, CPL) The Character of a Physical Law, MIT Press.

Feynman, R.P. (1997, SNP) Six Not so Easy Pieces. Einstein's Relativity Symmetry and Space-Time. Perseus Books, Cambridge, Mass.

Fraenkel, A., Bar Hillel, V., Levy, A. (1973, FST) Foundations of Set Theory, North Holland, Amsterdam.

Frege, G. (1964, BLA) The Basic Laws of Arithmetic. Parts translated and edited by M. Furth, University of California Press, Berkeley. -Complete German Edition: Grundgesetze der Arithmetik, Jena 1893 and 1903. Reprinted Wissenschaftliche Buchgesellschaft, Darmstadt 1962.

Frege, G. (1969, NGS) Nachgelassene Schriften, Band I, Hermes H. et al. (eds.), Meiner, Hamburg.

Frede, M., Patzig, G. (1988, AMP) Aristoteles "Metaphysik Z", 2 Vols, C.H. Beck, München.

Freedman, S.J. et al. (1972, ELH) Experimental Test of Local Hidden-Variable Theories, Phys. Rev. Lett. **28**, 938–941.

Friedman, M. (1983, FST) Foundations of Space Time Theories. Princeton Univ. Press.

Gabbay, D. and Guenther, F. (1984, HBP) Handbook of Philosophical Logic. Reidel, Dordrecht.

Galileo, G. (DNS) Dialogues Concerning Two New Sciences. Transl. by H. Crew and A. de Salvio, Dover, New York.

Galileo Galilei (DWS) Dialogue Concerning the Two Chief World Systems – Ptolemaic and Copernican. Transl. St. Drake, University of California Press, Berkeley, 1967.

Galles, D., Pearl, J. (1997, ACR) Axioms of Causal Relevance. Artificial Intelligence **97**, pp. 9–43.

Gamow, G. (1928, ZQA) Zur Quantentheorie des Atomkerns, Z. Physik **51**, p. 204 ff.

Gamow, G. (1965, MTP) Mr. Tompkins in Paperback, (Containing Mr. Tompkins in Wonderland and Mr. Tompkins explores the Atom). Cambridge University Press, Cambridge.

Gauss, C.F. (1828, Dgc) Disquisitiones generales circa superficies curvas, in: Werke Bd.4, Göttingen 1863–1903.

Gell-Mann, M., Hartle, J.B. (1994, TSA) Time Symmetry and Asymmetry in Quantum Mechanics and Quantum Cosmology. In: Halliwell, J.J. et al. (1994, POT), pp. 311–345.

Genz, H., Decker, R. (1991, SSB) Symmetrie und Symmetriebrechung in der Physik. Vieweg, Braunschweig.

Gisin, N. et al. (2000, OQN) Optical Tests of Quantum Nonlocality, Ann. Phys. (Leipzig), **9**, pp. 831–841.

Gödel, K. (1930, EMR) Einige metamathematische Resultate über Entscheidungsdefinitheit und Widerspruchsfreiheit. In: Anzeiger der Akademie der Wissenschaften in Wien **67**, pp. 214–215. Engl. Translation in: Kurt Gödel, Collected Works, Vol. I ed. by S. Feferman et al., Oxford 1986.

Gödel, K. (1931, FUS) Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. In: Monatshefte für Mathematik und Physik **38**, pp. 173–198. Engl. translation in: Ibid. Vol. I.

Gödel, K. (1940, CCH) The Consistency of the Axiom of Choice and the Generalized Continuum-Hypothesis with the Axioms of Set Theory, Annals of Mathematics Studies **3**, Princeton.

Gödel, K. (1944, RML) Russell's Mathematical Logic, in: Russell, B. (1944, PBR), pp. 125–153.

Goenner, H. (eds.) (1999, EWG) The Expanding Worlds of General Relativity. Birkhäuser, Boston.

Grasshoff, Th., May, M. (2001, CRg) Causal Regularities, in: Spohn, W. et al. (eds) (2001, CIC), pp. 85–114.

Gutmann, S. (1995,CPQ) Using classical probability to guarantee properties of infinite quantum systems, quant-ph/ 9506016.

Hadamard, J. (1898, SCO) Les surfaces à courbures opposées, Journal des Mathématiques pures et appliquées, 5th. Series, Vol. 4, pp. 27–73.

Hagiwara, K. et al. (2002, PDG) Particle Data Group, Physical Review D66, 010001.

Haldane, J.B.S. (1937, InM) The Inequality of Man, Penguin Book, Harmondsworth.

Halliwell, J.J., Perez-Mercader, J., Zurek, W.H. (1994, POT) Physical Origins of Time Asymmetry. Cambridge University Press, Cambridge.

Halliwell, J.J. (1994, QCT) Quantum Cosmology and Time Asymmetry. In: Halliwell, J.J. et al. (1994, POT), pp. 369–389.

Harper, W. and Smith, G. (1995, NNW) Newton's New Way of Inquiry. In: Jarrett, Leplin (eds.) The Creation of Ideas in Physics. Kluwer, Dordrecht.

Hausman, D.M. (1998, CAS) Causal Asymmetries, Cambridge University Press, Cambridge.

Hawking, S.W., Ellis, G.F.R. (1973, LSS) The Large Scale Structure of Space-Time. Cambridge University Press, Cambridge.

Hawking, S.W., Israel, W. (1979, GRE) General Relativity: An Einstein Centenary Survey. Cambridge University Press, Cambridge.

Hawking, S.W. (1980, TAG) Theoretical Advances in General Relativity. In: H. Woolf (ed.) (1980, SSP) pp. 145–152.

Hawking, S.W. (1980, ETP) Is the End in Sight for Theoretical Physics? Cambridge University Press, Cambridge.

Hawking, S.W., Israel, W. (1987, ECV) Einstein: A Centenary Volume, Cambridge University Press, Cambridge.

Hawking, S.W. (1988, BHT) A Brief History of Time, London, Bantam.

Healey, R. (1992, CRE) Causation, Robustness, and EPR, Philosphy of Science **59** (1992), pp. 282–292.

Hegerfeld, G.C., Ruijsenaars S.N. (1980, CLS) Remarks on Causality, Localisation, and Spreading of Wave Packets, Phys. Rev. D **22,** pp. 377–384.

Hegerfeld, G.C. (1998, ISQ) Instantaneous Spreading and Einstein Causality in Quantum Theory, Ann. Phys. (Leipzig) **7**, pp. 716–725.

Heintzmann, H., Mittelstaedt, P. (1968, PGB) Physikalische Gesetze in beschleunigten Bezugssystemen. In: Springer Tracts in Modern Physics **47**, pp. 185–225.

Heisenberg, W. (1967, ETE) Einführung in die einheitliche Theorie der Elementarteilchen, Hirzel Verlag, Stuttgart.

Hénon, M. (1976, TDM) A Two Dimensional-Map with Strange Attractor, Commun. Math. Phys. **50**, pp. 69–77.

Hilbert, D. (1922, NdM) Neubegründung der Mathematik, in: Abhandlungen aus dem Mathematischen Seminar der Hamburger Universität, Band 1, pp. 157–177.

Hilbert, D. (1962, GlG) Grundlagen der Geometrie, Teubner, Stuttgart.

Hintikka, K.J.J. (1966, AAn) An Analysis of Analyticity, in: Weingartner, P. (ed.) Deskription, Analytizität und Existenz, Pustet, Salzburg, pp. 193–214.

Hintikka, K.J.J. (1966, KVd) Kant Vindicated, in: Weingartner, P. (ed.) Deskription, Analytizität und Existenz, Pustet, Salzburg, pp. 234–253.

Holt, D.L., Holt, R.G. (1993, RND) Regularity in Nonlinear Dynamical Systems. Brit. J. Phil. Sci. **44**, pp. 711–727.

Howard, D. (1999, PCP) Point Coincidences and Pointer Coincidences: Einstein on the Invariant Content of Spacetime Theories. In: Goenner, H. et al. (1999, EWG) p. 463–500.

Hume, D. (1739, THN) A Treatise of Human Nature, Books I and II. Reprint: Clarendon Press, Oxford 1966.

Hume, D. (1740, THN) A Treatise of Human Nature, Book III. Reprint: Clarendon Press, Oxford 1966.

Hume, D. (1748, EHU) Enquiry into the Human Understanding, in: Enquiries, by David Hume, ed. L.A. Selby-Bigge, Oxford University Press, Oxford 1902 (1961).

Irvine, J.M. (1983, CLP) The Constancy of the Laws of Physics in the Light of Prehistoric Nuclear Reactors, Contemporary Physics **24(5)**, pp. 427–437.

Ismael, J. (1997, CPr) Curie's Principle, Synthese **110**, pp. 167–190.

Jammer, M. (1954, CSP) Concepts of Space. Harvard University Press, Cambridge, Mass.

Jammer, M., Stachel, J. (1979, MWA) If Maxwell had Worked between Ampère and Faraday. (Mimeographed Version, Dept. of Physics, Boston University).

Jauch, J.M. (1968, FQM) Foundations of Quantum Mechanics, Addison Wesley, Reading Mass.

Jauch, J.M., Piron, C. (1970, WQL) "What is Quantum Logic?" in: Quanta, Essays in Theoretical Physics dedicated to Gregor Wenzel, eds. P.G.O. Freyund, C.G. Goebel, Y. Nambu, University of Chicago Press.

Johannes Philoponos, (1899, DAM) De aeternitate mundi :contra Proclum. In: Ioannes Philoponus, De aeternitate mundi: Contra Proclum, ed. Hugo Rabe, Lipsiae 1899.

Kant, I. (1758, NLB) Neuer Lehrbegriff der Bewegung und Ruhe. In: Kant, I. (1905, GSW) Vol 2, pp. 12–25.

Kant, I. (1763, VBN) Versuch den Begriff der negativen Größe in die Weltweisheit einzuführen. In: Kant, I. (1905, GSW) Vol 2, pp. 165–204.

Kant, I. (1768, EGU) Von dem ersten Grunde des Unterschiedes der Gegenden im Raume. In: Kant, I. (1905, GSW) Vol 2, pp. 375–383.

Kant, I. (1783, PzM) Prolegomena zu einer jeden zukünftigen Metaphysik, (1905, GSW), Bd. IV.

Kant, I. (1786, MAN) Metaphysische Anfangsgründe der Naturwissenschaft. In: Kant (1905, GSW), Bd. IV.

Kant, I. (1787, KRV) Kritik der reinen Vernunft. In: Kant I. (1905, GSW) Bd. III and IV.

Kant, I. (1800, Lg) Logik. In: Kant (1905, GSW), Bd. IX.

Kant, I. (1905, GSW) Gesammelte Schriften, Akademie Ausgabe, Berlin.

Kant, I. (1929, CPR) Critque of Pure Reason, English Transl. of Kant, I. (1787, KRV) by N.K. Smith, MacMillan, London.

Kleinert, H. (1993, PDI) Pfadintegrale. BI Wissenschaftsverlag, Mannheim.

Kolmogoroff, A.N. (1933, GdW) Grundbegiffe der Wahrscheinlichkeitsrechnung. Ergebn. d. Math.2, Heft.3, Springer, Berlin. Reprint 1973.

Kolmogorov, A.N. (1950, FTP) Foundations of the Theory of Probability, Chelsea, New York, Engl. Translation of Kolmogoroff, A.N. (1933, GdW).

Kolmogorov, A.N. (1954, CCP) On Conservation of Conditionally-Periodic Motions for Small Change in Hamilton's Function, Dokl. Akad. Nauk, USSR **98**, p. 525 ff.

Kreisel, G. (1965, MLg) Mathematical Logic, in: Saaty, T.L. (ed.) Lectures in Modern Mathematics III, Wiley, New York, pp. 95–195.

Kreisel, G. (1967, MLW) Mathematical Logic: What has it done for the Philosophy of Mathematics, in: Schoeneman, R. (ed.) Bertrand Russell, Philosopher of the Century, Allen and Unwin, London, pp. 201–272.

Kreisel, G. (1974, NMT) A Notion of Mechanistic Theory, Synthese **29**, pp. 11–26.

Kreisel, G. (1980, KGB) Kurt Gödel, 28. 4.1906 -14.1.1978. Biographical Memoirs of Fellows of the Royal Society **26**, pp. 148–224; Correction **27**, p. 697 and **28**, p. 718.

Kreisel, G. (1990, RGC) Review of Gödel's Collected Works, Volume II, In: Notre Dame Journal of Formal Logic **31**, 4 (1990), pp. 602–64.

Kuratowski, K. (1966, Top) Topology, Vol. I and II, Academic Press, New York.

Kwiat et al. (1995, IBI) High-Visibiliy Interference in a Bell-Inequality Experiment for Energy and Time, Phys. Rev. Lett. **75**, 4337.

Kuhn, Th. (1978, BBT) Black-Body Theory and the Quantum Theory. 1894–1912, Oxford.

Lagrange, J.L. (1772, EPT) Essai sur le problème des trois corps. Paris.

Landau, L.D., Lifschitz E.M. (1966, ThP) Lehrbuch der Theoretischen Physik V, Statistische Physik, Akademie Verlag, Berlin.

Landau, L.D., Lifschitz, E.M. (1963, ThP) Lehrbuch der Theoretischen Physik, II, Aufbau-Verlag, Berlin

Laugwitz, D. (1960, DfG) Differentialgeometrie, B.G. Teubner, Stuttgart.

Laplace, P. (1814, EPr) Essai philosophique sur les probabilités. Courcier, Paris, Engl. Transl.: A Philosophical Essay on Probabilities. Dover, New York 1951.

Laplace, P.S. (1951, PEP) A Philosophical Essay on Probabilities, New York, Dover.

Laskar, J. (1994, LCS) Large Scale Chaos in the Solar System, Astronomy and Astrophysics **287**, L9–L12.

Laurikainen, K.V. (1988, BtA) Beyond the Atom, Springer, Heidelberg.

Lebowitz, J.L. (1994, TAB) Time's Arrow and Boltzmann's Entropy. In: Halliwell et al. (1994, POT), pp. 131–146.

Lee, T.D. (1988, SAW) Symmetries, Asymmetries and the World of Particles. University of Washington Press, Seattle.

Leibniz, G.W. (GPh) Die philosophischen Schriften von G.W. Leibniz. Ed. by C.I. Gerhardt, 7 vols., Berlin. 1875–90.

Leibniz, G.W. (Met) Discours de Métaphysique. In: Leibniz, G.W. (GPh), Vol. IV, p. 427ff.

Leibniz, G.W. (NE) Nouveaux essais sur l'entendement humain. In: Leibniz (GPh), Vol. V.

Leibniz, G.W. (1903, OFI) Opuscules et fragments inédits de Leibniz, ed. L. Couturat, Presses Universitaires de France, Paris 1903. Reprint : Olms, Hildesheim 1961.

Lewis, C.L. (1918, SSL) A Survey of Symbolic Logic, University of California Press, Berkeley.

Lewis, D. (1973, Caus) Causation, The Journal of Philosophy **70**, pp. 556–572.

Lewis, D. (1973, Ctf) Counterfactuals, Harvard University Press, Cambridge Mass.

Lewis, D. (2000, CIn) Causation as Influence, The Journal of Philosophy **97**, pp. 182–197.

Lie, M., Engel, S. (1888, TfG) Theorie der Transformationsgruppen,3 Bde. Leipzig, Bd. 3, Abt. 5.

Lighthill, J. (1986, RRF) The Recently Recognized Failure of Predictability in Newtonian Dynamics. In: Proceedings of the Royal Society London A **407**, pp. 35–50.

Linde, A. (1990, IQC) Inflation and Quantum Cosmology, Academic Press.

Lobachevsky, N.J. (1829, OFG) On the Foundations of Geometry, Kazan Bulletin.

Longair (1984, TCP) Theoretical Concepts in Physics. Cambridge University Press, Cambridge.

Lorentz, H.A., Einstein, A., Minkowski, H., Weyl, H. (1923, PRT) The Principle of Relativity. Dover, London.

Lorenz, E.H. (1963, DNF) Deterministic Nonperiodic Flow, J. Atmos. Science **20**, pp. 130–141.

Lorenzen, P. (1955, EOL) Einführung in die operative Logik und Mathematik, Springer, Berlin.

Lorenzen, P. (1980, MeM) Metamathematik, BI-Wissenschaftsverlag, Mannheim.

Lüders, G. (1951, ZMP) Über die Zustandsänderung im Meßprozeß, Ann. Phys. 6. Folge, Vol. 8, pp. 322–328.

Mach, E. (1933, MEC) Die Mechanik in ihrer Entwicklung. Leipzig (9th edition).

Mainzer, K. (1980, GdG) Geschichte der Geometrie, BI-Wissenschaftsverlag, Mannheim

Mainzer, K. (1984, OpM) Mathematik, operative in: Enzyklopädie Philosophie und Wissenschaftstheorie 2, BI-Wissenschaftsberlag, Mannheim, pp. 806–809.

March, A. (1957, NDM) Das neue Denken der modernen Physik. Rowohlt, Hamburg.

March, A. (1960, PEG) Die Physikalische Erkenntnis und ihre Grenzen, Vieweg, Braunschweig.

Margenau, H. (1944, Epr) The Exclusion Principle and its Philosophical Importance, Philosophy of Science **11**, pp. 187–208.

Mathieu, V. (1989, KOP) Kants Opus postumum, Klostermann, Frankfurt.

Maxwell, J.C. (1991, MaM) Matter and Motion. Dover, New York (Based on the Edition by J. Larmor of 1920).

Meessen, A. (2000, STQ) Spacetime Quantisation, Elementary Particles and Cosmology, Foundations of Physics **29**, pp. 281–316.

Mehra, J. (ed.) (1973, PCN) The Physicist's Concept of Nature, Reidel, Dordrecht, Holland.

Meschkowski, H. (1978, PNM) Problemgeschichte der neueren Mathematik, B.I. Wissenschaftsverlag, Mannheim.

Miles, J. (1984, RMS) Resonant Motion of a Spherical Pendulum, Physica **11D**, pp. 309–323.

Miller, D. (1974, CFT) On the Comparison of False Theories by their Bases. British Journal for the Philosophy of Science **25**, pp. 178–188.

Misner, Ch.W., Thorne, K.S., Wheeler, J.A. (1973, Grav) Gravitation. Freeman, New York.

Mittelstaedt, P. (1978, QLg) Quantum Logic, Reidel, Dordrecht.

Mittelstaedt, P. and E.-W. Stachow (1978, PEM) The Principle of Excluded Middle in Quantum Logic, Journal of Philosophical Logic, **7**, pp. 181–208.

Mittelstaedt, P. (1984, CNT) Constituting, Naming, and Identity in Quantum Logic. In: Recent Developments in Quantum Logic, B.I. Wissenschaftsverlag, Mannheim, pp. 215–234.

Mittelstaedt, P. (1986, STP) Sprache und Realität in der modernen Physik, BI-Wissenschaftsverlag, Mannheim.

Mittelstaedt, P. (1989, PMP) Philosophische Probleme der Modernen Physik, 7. ed, B.I. Wissenschaftsverlag, Mannheim. Engl. Trans. Philosophical Problems of Modern Physics, D. Reidel Publ. Co., Dordrecht, 1976.

Mittelstaedt, P. (1989, ZBP) Der Zeitbegriff in der Physik. BI-Wissenschaftsverlag, Mannheim.

Mittelstaedt, P. (1990, OMI) The Objectification in the Measuring Process and the Many-Worlds Interpretation, in: Symposium on the Foundations of Modern Physics 1990, World Scientific, Singapore, pp. 261–279.

Mittelstaedt, P. (1993, MPI) The Measuring Process and the Interpretation of Quantum Mechanics, Int. Journ. of Theor. Phys. **32**, pp. 1763–1775.

Mittelstaedt, P. (1994, OKP) The Constitution of Objects in Kant's Philosophy and in Modern Physics. In: P. Parrini, ed., Kant and Contemporary Epistemology, Kluwer, pp. 115–129.

Mittelstaedt, P. (1995, OQM) Constitution of Objects in Classical Mechanics and in Quantum Mechanics. Int. Journal of Theor. Physics, **34**, pp. 1615–1626.

Mittelstaedt, P. (1995, KLM) Klassische Mechanik. 2 ed., BI-Wissenschaftsverlag, Mannheim.

Mittelstaedt, P. (1997, QPT) Is Quantum Mechanics a probabilistic Theory? in: R.S. Cohen, et al. (eds.) Potentiality, Entanglement and Passion-at-a-Distance, Kluwer Academic Publishers.

Mittelstaedt, P. (1997, ESL) The Emergence of Statistical Laws in Quantum Mechanics. In: M. Ferrero, A. van der Merwe (eds.): New Developments on Fundamental Problems in Quantum Physics, Kluwer Acad. Publ., pp. 265–274.

Mittelstaedt, P. (1998, IQM) The Interpretation of Quantum Mechanics and the Measurement Process, Cambridge University Press, Cambridge.

Mittelstaedt, P., Nimtz, G. (eds.) (1998, SLV) Superluminal(?) Velocities. Annalen der Physik 7–8 (1998).

Mittelstaedt, P. (2000, WSS) What if there are superluminal signals? Eur. Phys. J.B. **13** pp. 353–355.

Möller, C. (1952, ThR) The Theory of Relativity. Oxford University Press.

Moser, J. (1967, CSE) Convergent Series Expansions of Quasi-Periodic Motions, Math. Ann. **189**, pp. 163 ff.

Moser J. (1978, SSS) Is the Solar system Stable? Mathematical Intelligencer **1**, pp. 65–71.

Newton, I. (Princ) Mathematical Principles of Natural Philosophy. Ed. F. Cajori, Univ. of California Press, Berkeley, 1934.

Newton, I. (Grav) De Gravitatione. In: Hall, A.R., Hall, M.B. Unpublished Scientific Papers of Isaac Newton. Cambridge U.P., 1962, pp. 89–156.

Newton, I. (Opt) Optics. Dover, New York, 1952.

Niiniluoto, I. (1998, VTP) Verisimilitude: The Third Period. British Journal for the Philosophy of Science **49**, pp. 1–29.

Nimtz, G. (2003, SLT) On Superluminal Tunneling, Progress in Quantum Electronics **27**, pp. 417–450.

Noll, W. (1967, STS) Space-Time Structures in Classical Mechanics. In: Bunge, M. (1967, DSF) pp. 28–34.

Norton, J. (1989, CCE) Coordinates and Covariance: Einstein's View of Space-Time and the Modern View. In: Foundations of Physics **19** (1989) pp. 1215–1263.

Norton, J. (1993, GCF) General Covariance and the Foundations of General Relativity: Eight Decades of Dispute. In: Reports on Progress in Physics **56**, pp. 791–858.

Ockham, W.v. (1957, SLg) Summa Logicae, ed. Ph. Boehner, Pars Prima, New York.

Ornstein, D.S., Weiss, B. (1991, SPC) Statistical Properties of Chaotic Systems, Bulletin of the American Mathematical Society **24**, pp. 11–116.

Pagels, H. (1983, CSC) The Cosmic Code, Bantam, New York.

Paris, J. and L. Harrington (1977, MIP) A Mathematical Incompleteness in Peano Arithmetic. In: Barwise, Handbook of Mathematical Logic, Amsterdam, North Holland, pp. 1133–1142.

Parry, W.T. (1933, AAI) Ein Axiomsystem für eine neue Art von Implikation (Analytische Implikation), Ergebnisse eines mathematischen Kolloquiums **4**, pp. 5–6.

Peano, G. (1889, PAP) The Principles of Arithmetic Presented by a New Method. In: Van Heijenoort, J. (ed.), From Frege to Gödel. Harvard University Press, Cambridge, Mass. pp. 83–97.

Pearl, J. (2000, CMR) Causality, Models, Reasoning and Inference. Cambridge University Press, Cambridge.

Peirce, Ch.S., (1960, CPC) Collected Papers of Ch.S. Peirce, ed. by Ch. Hartshorne and P. Weiss, Harvard U. P., Cambridge Mass.

Penrose, R. (1979, STA) Singularities and Time-Asymmetry. In: Hawking, Israel (1979, GRE), pp. 581–638.

Penrose, R. (1997, LSH) The Large, the Small and the Human Mind, Cambridge University Press.

Penrose, R. (2001, GRQ) On Gravity's Role in Quantum State Reduction. In: Callender, C. and Huggett, N. (2001, PMP).

Pesin, Ya.B. (1977, CLE) Characteristic Ljapunov Exponents and Smooth Ergodic Theory. In: Russian Math. Surveys **32**(4), pp. 55–114.

Petley, B.W. (1985, FPC) The Fundamental Physical Constants and the Frontier of Measurement, Adam Hilger, Bristol.

Petley, B.W. (1999, FdK) Fundamentalkonstanten. In: Lexikon der Physik, Vol. II. Spektrum Heidelberg, pp. 429–434.

Piron, C. (1964, AQu) Axiomatique Quantique, Helv. Phys. Acta **37**, p. 439.

Piron, C. (1976, FQP) Foundations of Quantum Physics, W.A. Benjamin, Reading Mass.

Pitowsky, I. (1989, QPL) Quantum Probability – Quantum Logic, Lecture Notes in Physics, Vol. 321, Springer, Berlin. 1994.

Pitowsky, I. (1994, CPE) George Boole's "Conditions of Possible Experience" and the Quantum Puzzle, Brit. Journ. Philos. Sci. **45**, pp. 95–125.

Planck, M. (1913, THR) The Theory of Heat Radiation, (transl. By M. Masius), Dover, New York 1959.

Poincaré, H. (1892, MNM) Les Methodes Nouvelles de la Méchanique Celeste. Vol. I, Paris.

Poincaré, H. (1952, ScH) Science and Hypothesis. Dover, New York.

Poincaré, H. (1958, VSc) Values of Science. Dover, New York.

Popper, K.R. (1959, LSD) The Logic of Scientific Discovery, Hutchinson, London.

Popper, K.R. (1959, PIP) The Propensity Interpretation of Probability, British Journal for the Philosophy of Science **10**, pp. 25–42.

Popper, K.R. (1963, CaR) Conjectures and Refutations, Routledge and Kegan Paul, London.

Popper, K.R. (1965, CaC) Of Clouds and Clocks. In: Popper: Objective Knowledge. Oxford 1972, pp. 206–255.

Popper, K.R. (1969, EKS) Epistemology Without a Knowing Subject. In: Rootselaar, B. (ed.) Logic, Methodology and the Philosophy of Science, Vol. 3, North Holland, Amsterdam, pp. 333–373.

Popper, K.R. (1972, OKn) Objective Knowledge: An Evolutionary Approach, Clarendon Press, Oxford.

Popper, K.R. (1982, OUn) The Open Universe, Rowman and Littlefield, Totowa, New Jersey.

Popper, K.R. (1983, RAS) Realism and the Aim of Science, Huchinson, London.

Popper, K.R., Eccles, J. (1985, SBr) The Self and its Brain, Springer, Berlin.

Post, E.J. (1967, PNG) On the Physical Necessity for General Covariance in Electromagnetic Theory. In: Bunge, M. (ed.) (1967, DSF).

Priest, G. (1987, ICd) In Contradiction. M. Nijhoff, The Hague.

Prior, A. (1957, TMd) Time and Modality. Oxford University Press, Oxford.

Prior, A. (1967, PPF) Past, Presence and Future. Oxford University Press, Oxford.

Prigogine, I. (1993, TDC) Time, Dynamics and Chaos: Integrating Poincaré's Non-Integrable Systems. In: Holte, J. (ed.) Nobel Conference XXVI. Chaos: The New Science. London, University Press of America.

Prigogine, I. (1995, GCh) Die Gesetze des Chaos. Campus Verlag, Frankfurt.

Prigogine, I., Stengers, I. (1993, PZt) Das Paradox der Zeit. Zeit, Chaos und Quanten, München, Piper.

Putnam, H. (1969, ILE) Is Logic Empirical? in: Boston Studies in the Philosophy of Science, Vol. V, D. Reidel Publ. Co. Dordrecht, Holland.

Quine, W.v.O. (1961, LPV) From a Logical Point of View, Harvard University Press, Cambridge, Mass.

Quine, W.v.O. (1969, MLg) Mathematical Logic. Harvard University Press, Cambridge, Mass.

Quine, W.v.O. (1970, PLg) Philosophy of Logic, Prentice-Hall, Englewood Cliffs.

Rauch, H. (1988, NIQ) Neutron Interferometric Tests of Quantum Mechanics, Helv. Phys. Acta **61**, p. 589.

Rees, M.J. (1983, LNR) Large Numbers and Ratios in Astrophysical Cosmology. Phil. Trans. Royal Soc., London **A 310**, pp. 311–322.

Rees, M.J. (2002, NCT) Numerical Coincidence and "Tuning" in Cosmology. In: Wickramasinghe et al. (ed), Fred Hoyle's Universe, Kluver, Dordrecht, pp. 95–108.

Redhead, M. (1987, INR) Incompleteness, Nonlocality and Realism, Oxford University Press, Oxford.

Rescher, N. (1967, PhL) The Philosophy of Leibniz, Prentice-Hall, Englewood Cliffs.

Rescher, N. (1969, MVL) Many Valued Logic, McGraw Hill, New York.

Resnik, R. and Halliday, D. (1985, BCR) Basic Concepts in Relativity and Early Quantum Theory. J. Wiley, New York.

Richter. H. (1956, WTh) Wahrscheinlichkeitstheorie, Springer, Berlin-Heidelberg.

Riemann, B. (1854, HyG) Über die Hypothesen welche der Geometrie zugrunde liegen, in: Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen, 1867. Suhrkamp Verlag, Frankfurt/M.

Rindler, W. (1977, ESR) Essential Relativity – Special, General and Cosmological. 2nd ed., Van Nostrand, New York.

Rohrlich, F. (1965, CCP) Classical Charged Particles. Addison Wesley, Mass.

Rosenthal-Schneider, I. (1949, PAE) Presuppositions and Anticipations in Einstein's Physics. In: Einstein, A. (1949, EPS) pp. 131–146.

Rosenthal-Schneider, I. (1980, RST) Reality and Scientific Truth: Discussions with Einstein, von Laue, and Planck. Wayne State University Press, Detroit.

Ruelle, D. (1980, StA) Strange Attractors, Mathematical Intelligencer **2**, pp. 126–137.

Ruelle, D. (1990, DCh) Deterministic Chaos: The Science and the Fiction. Proc. Roy. Soc. London, **A 427**, pp. 241–248.

Ruelle, D. (1991, CCh) Chance and Chaos, Princeton University Press, Princeton.

Russell, B. (1919, IMP) Introduction to Mathematical Philosophy, Allen and Unwin, London.

Russell, B. (1944, PBR) The Philosophy of Bertrand Russell, ed. P.A. Schilpp, 3rd ed. 1951, Tudor, New York.

Rynasiewicz, R. (1999, KAC) Kretschmann's Analysis of Covariance and Relativity Principles. In: Goenner, H. et al. (1999, EWG) pp. 431–462.

Salam, A., Wigner, E.P. (eds.) (1972, AQT) Aspects of Quantum Theory. Cambridge University Press, Cambridge.

Sakurai, J.J. (1994, MQM) Modern Quantum Mechanics, 2nd ed., Addison Wesley, Reading, Mass.

Sarton, G. (1966, HSc) A History of Science. Harvard University Press, Cambridge, Mass.

Scheibe, E. (1997, RPT) Die Reduktion physikalischer Theorien, Vol. 1, Springer Verlag, Heidelberg, Berlin.

Scheibe, E. (1999, RPT) Die Reduktion physikalischer Theorien, Vol. 2, Springer Verlag, Heidelberg, Berlin.

Schelb, U. (1998, STA) On the Role of a Limiting Velocity in Constructive Spacetime Axiomatics, Ann. Phys. (Leipzig), **7**, pp. 748–755.

Schlieder, S. (1968, ZML) Einige Bemerkungen zu Zustandsänderungen von rel. quantenmechanischen Systemen durch Messungen und zur Lokalitätsforderung, Commun. Math. Phys. **7**, pp. 305–331.

Schlieder, S. (1971, KRQ) Zum kausalen Verhalten eines relativistischen quantenmechanischen Systems, in P. Dürr, ed. Quanten und Felder, Vieweg, Braunschweig, pp. 145–160.

Schmutzer, E. (1996, RTA) Relativitätstheorie aktuell. Teubner, Leipzig.

Schrödinger, E. (1944, WLf) What is Life? Cambridge University Press, Cambridge.

Schrödinger, E. (1954, STS) Spacetime Structure. Cambridge University Press, Cambridge.

Schrödinger, E. (1961, WNG) Was ist ein Naturgesetz? Oldenbourg, Munich.

Schulman, L.S. (1994, TSC) Time-Symmetric Cosmology and Definite Quantum Measurements. In: Halliwell et al. (1994, POT), pp. 299–308.

Schurz, G., Weingartner, P. (1987, VDR) Versimilitude Defined by Relevant Consequence Elements, in: Kuipers, Th. (ed.) What is Closer to the Truth, Rodopi, Amsterdam, pp. 47–77.

Schurz, G. (2001, CAI) Causal Asymmetry, Independent versus Dependent Variables, and the Direction of Time. In: Spohn, W. et al. (2001, CIC), pp. 47–67.

Schuster, H.G. (1989, DCh) Deterministic Chaos, VCH, Weinheim.

Schwarzschild, B. (1999, TEO): Two Experiments Observe Explicit Violation of Time-Reversal Symmetry. In: Physics Today, Feb. 1999, pp. 19–20.

Sexl, R., Urbantke, H. (1992, RGT) Relativität, Gruppen, Teilchen. Springer-Verlag, Wien.

Sexl, R., Urbantke, H. (2002, GVK) Gravitation und Kosmologie. Eine Einführung in die Allgemeine Relativitätstheorie. Spectrum, Heidelberg – Berlin.

Shapiro, I.I. (1980, ECP) Experimental Challenges Posed by the General Theory of Relativity. In: Woolf (ed.) (1980, SSP) pp. 115–135.

Sklar, L. (1977, CTT) What Might be Right about the Causal Theory of Time. In: Synthese **35** (1977), pp. 155–171.

Skolem, Th. (1970, MLg) Über die mathematische Logik. In: Skolem, Th., Selected Works J.E. Fenstad (ed.) Universitetsforlaget, Oslo 1970, pp. 187–206.

Smith, K. (1966, PDH) The Philosophy of David Hume, London (1941, 1966).

Spinoza, B. (1677, Eth) Ethica, Ordine Geometrico demonstrata. Jan Rieuwertsz, Amsterdam.

Spohn, W. et al. (eds.) (2001, CIC) Current Issues in Causation, Mentis, Paderborn.

Stachow, E.-W., (1980, LFQ) Logical Foundations of Quantum Mechanics, Int. Journ. of Theoretical Physics, **19**, pp. 251–304.

Stachow, E.-W., (1984, QLI) Structures of a Quantum Language for Individual Systems. In: Recent Developments in Quantum Logic, (eds.) P. Mittelstaedt and E.-W. Stachow, BI-Wissenschaftsverlag, Mannheim.

Stöltzner, M., Weingartner, P. (eds.) (2005, FTK) Formale Teleologie und Kausalität, Mentis, Paderborn.

Strohmeyer, I. (1995, QTP) Quantentheorie und Transzendentalphilosophie, Spektrum, Heidelberg.

Sudarshan, E.C.G., Mukunda, N. (1974, CDM) Classical Dynamics: A Modern Perspective, John Wiley & Sons, New York.

Suppes, P. (1970, PTC) A Probabilistic Theory of Causality, North-Holland, Amsterdam.

Suppes, P. (2002, RIS) Representation and Invariance of Scientific Structures, CSLJ, Stanford.

Tarski, A. (1944, SCT) The Semantic Concept of Truth and the Foundations of Semantics, Philosophy and Phenomenological Research **4**, pp. 341–376.

Tarski, A. (1951, DME) A Decision Method for Elementary Algebra and Geometry, University of California Press, Berkeley.

Tarski, A. (1956, LSM) Logic, Semantics, Metamathematics. Oxford University Press, Oxford.

Tarski, A. (1956, CLC) The Concept of Logical Consequence, in: Tarski, A. (1956, LSM), Ch. XVI.

Tarski, A., Mostowski, A., Robinson, R. (1968, UDT) Undecidable Theories. North-Holland, Amsterdam.

Tetens, H. (1994, AAp) Arithmetik: ein Apriori der Erfahrung? in: Dialektik 1994/3, Naturalismus in der Philosophie der Mathematik? Hrsg. Brigitte Falkenburg, pp. 125–146.

Thirring, W. (1979, CMP) A Course in Mathematical Physics. Vol. 2, Transl. by E.M. Harrell, Springer, New York.

Tichy, P. (1974, PDV) On Popper's Definitions of Verisimilitude, British Journal of the Philosophy of Science **25**, pp. 155–165.

T'Hooft, G. (1995, PVW) In: J. Hilgenvoord (ed.), Physics and our View of the World, Cambridge University Press, Cambridge.

Thomas Aquinas (CAP) Commentary on Aristotle's Physics. Transl. by R.J. Blackwell et al., Intr. by V.J. Bourke. Yale University Press, New Haven, 1963.

Thomas Aquinas (STh) Summa Theologica. Transl. by Fathers of the English Dominican Province (Christian Classics, Westminster, Maryland, 1948, 1981).

Thorne, K. (1994, BHW) Black Holes and Time Warps, W.W. Norton & Co. Inc.

Torretti, R. (1983 RaG) Relativity and Geometry. Pergamon, New York.

van Benthem, J. (1984, FCL) Foundations of Conditional Logic, Journal of Philosophical Logic **13**, pp. 303–349.

van Benthem, J. (1991, LgT) The Logic of Time. Kluwer, Dordrecht.

van Fraassen, B. (1989, LaS) Laws and Symmetries, Oxford University Press, New York.

Vilenkin, A. (1982, CUN) Creation of Universes from Nothing, Physics Letters 117B, pp. 25–28.

von Helmholtz, H. (1879, TdW) Die Tatsachen in der Wahrnehmung, August Hirschwald, Berlin.

von Helmholtz, H. (1868, TdG) Über die Tatsachen, die der Geometrie zu Grunde liegen, Nachrichten der Königlichen Gesellschaft der Wissenschaften und der Georg-Augusts-Universität, Nr.9, pp. 193–221.

von Helmholtz, H. (1984, UBA) Über den Ursprung und die Bedeutung der geometrischen Axiome, Vorträge und Reden, II. Braunschweig.

von Kutschera, F. (1993, Caus) Causation. Journal of Philosophical Logic, **22**, pp. 563–588.

von Mises, R. (1931, WAS) Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik, Wien 1931.

von Neumann, J. (1932, MGQ) Mathematische Grundlagen der Quantenmechanik, Springer Verlag, Berlin.

von Neumann, J. (1938, IDP) On infinite direct products, Compositio Mathematica **6**, pp. 1–77.

Weierstrass, K. (1993, BKW) Briefwechsel zwischen Karl Weierstrass und Sofia Kowalevskaya, (ed.) R. Bölling, Akademie Verlag, Berlin.

Weihs et al. (1998, VBI) Violation of Bell's Inequality under Strict Einstein Locality Conditions, Phys. Rev. Lett. **81,** pp. 5039–5043.

Weinberg, St. (1977, FTM) The First Three Minutes, London, Deutsch.

Weinberg, St. (1983, OTP) Overview of theoretical prospects for understanding the values of fundamental constants. Phil. Trans. Royal Soc. London **A310**, pp. 249–252.

Weinberg, St. (1987, TFL) Towards the Final Laws of Physics. In: Elementary Particles and the Laws of Physics. The 1986 Dirac Memorial Lectures, Cambridge University Press, Cambridge, pp. 61–110.

Weinberg, St. (1989, CCP) The Cosmological Constant Problem. Reviews of Modern Physics **61**, pp. 1–23.

Weinberg, St. (1993, DFT) Dreams of a Final Theory, Pantheon Books, New York.

Weingartner, P. (1968, MLT) Modal Logics with two Kinds of Possibility and Necessity, Notre Dame Journal of Formal Logic, pp. 97–159.

Weingartner, P. (1975, FAM) A Finite Approximation to Models of Set Theory, Studia Logica **34**, pp. 45–58.

Weingartner, P. (1976, WTh) Wissenschaftstheorie II, 1. Grundprobleme der Logik und Mathematik. Frommann-Holzboog, Stuttgart.

Weingartner, P. (1978, WThI) Wissenschaftstheorie I. Einführung in die Hauptprobleme, Frommann-Holzboog, Stuttgart.

Weingartner, P. (1982, DLM) On the Demarcation between Logic and Mathematics, The Monist **65**, pp. 38–51.

Weingartner, P. (1983, IMS) The Ideal of Mathematization of All Sciences and of "More Geometrico" in Descartes and Leibniz. In: Shea, W.R. (ed.), Nature Mathematized, Dordrecht, Holland, pp. 151–195.

Weingartner, P. (1985, SRC) A simple Relevance Criterion for Natural Language and its Semantics, in: Dorn, G., Weingartner, P. (eds.), Foundations of Logic and Linguistics: Problems and their Solutions, Plenum Press, New York, pp. 563–575.

Weingartner, P., Schurz, G. (1986, PSS) Paradoxes Solved by Simple Relevance Criteria, Logique et Analyse **113**, pp. 3–40.

Weingartner, P., Schmetterer, L. (1987, GRb) Gödel Remembered. Bibliopolis, Napoli.

Weingartner, P. (1988, RCC) Remarks on the Consequence Class of Theories, in: Scheibe, E. (ed.) The Role of Experience in Science, de Gruyter, Berlin, pp. 161–180.

Weingartner, P., Schurz, G. (eds.) (1996, LPL) Law and Prediction in the Light of Chaos Research. Lecture Notes in Physics 473, Springer, Berlin.

Weingartner, P. (1996, UWT) Under what Transformations are Laws Invariant? In: Weingartner, Schurz (1996, LPL), pp. 47–88.

Weingartner, P. (1997, CLN) Can the Laws of Nature (of Physics) be Complete? in: Dalla Chiara, M.L. et al. (eds.) Logic and Scientific Method, Kluwer, Dordrecht, pp. 429–446.

Weingartner, P. (1999, SLG) Are Statistical Laws Genuine Laws? A Concern of Poincaré and Boltzmann. In: Philosophia Scientiae **3**(2) 1998/99, pp. 215–236.

Weingartner, P. (2000, RFC) Reasons for Filtering Classical Logic, in: Batens, D. et al. (eds.) Frontiers in Paraconsistent Logics, Research Studies Press, London, pp. 315–327.

Weingartner, P. (2000, BQT) Basic Questions on Truth, Kluwer, Dordrecht.

Weingartner, P. (2001, ALO) Applications of Logic outside Logic and Mathematics, in: Stelzner, W., Stöckler, M. (eds.) Zwischen traditioneller und moderner Logik, Mentis, Paderborn, pp. 53–64.

Weingartner, P. (2004, RSL) Reasons from Science for Limiting Classical Logic. In: P. Weingartner (ed.) (2004, Alg) pp. 233–248.

Weingartner, P. (ed.) (2004, Alg) Alternative Logics. Do Sciences Need Them? Springer, Berlin.

Wess, J. (1989, CPT) The CPT-Theorem and its Significance for Fundamental Physics. In: Nuclear Physics B (Proc. Suppl.) **8**, p. 461 ff.

Weyl, H. (1923, MAR) Mathematische Analyse des Raumproblems, Springer Verlag, Berlin (1966), Philosophie der Mathematik und der Naturwissenschaften, Oldenburg.

Weyl, H. (1949, PMN) Philosophy of Mathematics and Natural Sciences, Princeton University Press.

Weyl, H. (1966, PMN) Philosophie der Mathematik und der Naturwissenschaften, Oldenburg.

Wheeler, J.A., Feynman, R.P. (1945, IAR) Interaction with the absorber as mechanics of radiation, Rev. Mod. Phys. **17**, pp. 157–181.

Wheeler, J.A., Feynman, R.P. (1949, DIA) Classical Electrodynamics in Terms of Direct Interparticle Action, Rev. Mod. Phys. **21**, pp. 425–434.

Wheeler, J.A. (1973, FRM) From Relativity to Mutability. In: Mehra, J. (ed.) (1973, PCN) pp. 202–247.

Wheeler, J.A. (1983, RLL) On Recognizing "Law without Law". In: American Journal of Physics **51**, pp. 398–404.

Wheeler, J.A. (1994, TTd) Time Today. In: Halliwell et al. (1994, POT), pp. 1–29.

Whitrow, G.J. (1978, IIP) On the Impossibility of an Infinite Past, British Journal of the Philosophy of Science, **29**, pp. 39–45.

Whyte, L.L. (1970, PCP) Pierre Curie's Principle of One-Way Process, Studium Generale, **23**, pp. 525–532. Reprinted in: Fraser, J.T. et al. (eds.), The Study of Time, Springer, Berlin 1972, pp. 140–147.

Wieland, W. (1970, APh) Die Aristotelische Physik, Vandenhoeck und Ruprecht, Göttingen.

Wigner, E.P. (1932, OZQ) Über die Operation der Zeitumkehr in der Quantenmechanik. Göttinger Nachrichten **31**, pp. 546–559.

Wigner, E.P. (1967, SRf) Symmetries and Reflections. Scientific Essays of Eugene P. Wigner. Ed. by Moore, W.J. and Scriven, M., Indiana U.P., Bloomington.

Wigner, E.P. (1972, TEU) On the Time Energy Uncertainty Relation. In: Salam, A., Wigner, E.P. (1972, AQT), pp. 237–247.

Wittgenstein, L. (1960, PhI) Philosophische Untersuchungen, Schriften 1, Suhrkamp, Frankfurt/M. English Transl. by G.E.M. Anscombe, Blackwell, Oxford 1958.

Wittgenstein, L. (1960, TLP) Tractatus Logico Philosophicus, Schriften 1, Suhrkamp, Frankfurt/M. (Annalen der Naturphilosophie **14**, 1921) English Transl. by F.P. Ramsey, Routledge & Kegan Paul Ltd., London, 1922.

Wittgenstein, L. (1974, GdM) Bemerkungen über die Grundlagen der Mathematik, Schriften 6, Suhrkamp, Frankfurt/M., English Transl. by G.H. Wright, R. Rhees, and G.E.M. Anscombe, Blackwell, Oxford 1956.

Wolters, G. (1987, MER) Mach I, Mach II, Einstein und die Relativitätstheorie. Eine Fälschung und ihre Folgen. De Gruyter, Berlin.

Woodward (2001, LEB) Law and Explanation in Biology: Invariance is the Kind of Stability that Matters. In: Philosophy of Science **68**, pp. 1–20.

Woody, D.R., Richards, P.L. (1979, CBR) Spectrum of the Cosmic Background Radiation, Phys. Rev. Lett. **42**, pp. 925–29.

Woolf, H. (ed.) (1980, SSP) Some Strangeness in the Proportion. A Centennial Symposium to Celebrate the Achievements of Albert Einstein. Addison Wesley, Reading, Massachusetts.

Zeeman, E.C. (1964, CLG) Causality implies the Lorentz group, Journ. of Math. Physics **5**, pp. 490–493.

Zermelo, E. (1896, SDM) Über einen Satz der Dynamik und die mechanische Wärmetheorie. Wiedemanns Annalen **57**, 485 ff.

Zermelo, E. (1896, MEI) Über die mechanische Erklärung irreversibler Vorgänge. Wiedemanns Annalen **59**. 793 ff.

Zurek, W.H. (1994, PSS) Preferred Sets of States, Predictability, Classicality, and Environment-Induced Decoherence. In: Halliwell et al. (eds.) (1994, POT).

Index